

# Solving Two-Player Zero-Sum Repeated Bayesian Games

Lichun Li, Cedric Langbort and Jeff Shamma

## Abstract

This paper studies two-player zero-sum repeated Bayesian games in which every player has a private type that is unknown to the other player, and the initial probability of the type of every player is publicly known. The types of players are independently chosen according to the initial probabilities, and are kept the same all through the game. At every stage, players simultaneously choose actions, and announce their actions publicly. For finite horizon cases, an explicit linear program is provided to compute players' security strategies. Moreover, based on the existing results in [1], this paper shows that a player's sufficient statistics, which is independent of the strategy of the other player, consists of the belief over the player's own type, the regret with respect to the other player's type, and the stage. Explicit linear programs are provided to compute the initial regrets, and the security strategies that only depends on the sufficient statistics. For discounted cases, following the same idea in the finite horizon, this paper shows that a player's sufficient statistics consists of the belief of the player's own type and the anti-discounted regret with respect to the other player's type. Besides, an approximated security strategy depending on the sufficient statistics is provided, and an explicit linear program to compute the approximated security strategy is given. This paper also obtains a bound on the performance difference between the approximated security strategy and the security strategy.

## I. INTRODUCTION

In many games, players don't have complete information on the other players. Take cyber security problem as an example. Defenders may not know the payoffs, attacking measures, coordination methods of attackers. This paper studies a class of such games in which every player has a private type that is unknown to the other players. Although unknown to the others, the types of the players satisfy an initial probability distribution, which is a common knowledge. These games are called Bayesian games.

In most prior work in the literature of Bayesian games, one's belief on the types of the other players plays an important role in figuring out the Nash equilibrium or the perfect Bayesian equilibrium. Generally speaking, a player's belief of the other players' type depends on the other players' strategies. A special class of Bayesian games in which the common information based beliefs are strategy independent was considered in [2]. Because of the decoupling between the strategies and the beliefs, a backward induction algorithm was given to find Nash equilibria of the game. The cases when the beliefs of the players are strategy dependent were considered in [3], [4]. Both papers used perfect Bayesian equilibrium as their solution concept. Perfect Bayesian equilibrium consists of a strategy profile and a belief system such that the strategies are sequential rational given the belief system and the belief system is consistent given the strategy profile [5]. Based on the common information based belief system, [3] studied common information based perfect Bayesian equilibrium, and [4] studied structured perfect Bayesian equilibrium. Backward recursive formulas were given in both papers to find the corresponding perfect Bayesian equilibrium.

Bayesian games are also called games with incomplete information, and were studied in [6], [1], [7]. This prior work mainly studied two-player zero-sum repeated Bayesian games, which are the most closely related work of this paper. In two-player zero-sum repeated Bayesian games, minmax value, maxmin value, and game value are examined. Minmax value and maxmin value are also called the security level of the minimizer and the maximizer, respectively [8]. The strategy that guarantees the security level of the maximizer is called the security strategy of the maximizer, and the strategy assuring the security level of the minimizer is called the security strategy of the minimizer. When minmax value equals to maxmin value, we say the game has a value. In other words, there exists a Nash equilibrium, which is the security strategy pair. It was shown that the Nash equilibrium exists for both finite horizon and discounted two-player zero-sum repeated Bayesian games, but may not exist in infinite horizon average payoff two-player zero-sum repeated Bayesian games [6]. Later, [1], [9] provided backward recursive formulas to compute the game value for finite horizon case and discounted case. In the same paper, dual games of two-player zero-sum repeated Bayesian games were also studied. Besides backward recursive formulas, it was shown that the security strategy of a player in the dual game with special initial parameters is also the player's security strategy in the primal game, and that the security strategy only depends on the sufficient statistics consisting of the belief on the player's own type, a real vector with the same size as that of the other player's type set, and the stage if this is a finite horizon game. The physical meaning of the real vector in the primal game was still not clear in [1], [9]. Independently, [10] also showed that the uninformed player's sufficient statistics in an asymmetric repeated vector payoff game is also a real vector on the other player's type. Since the sufficient statistics only depends on a player's own strategy and signals available to the player [1], we call it self-dependent sufficient statistics.

This paper adopts the same game model as the one used in [6], [9], [7], [1], and has particular interests in developing prescriptive methods for players, i.e. computing the security strategies of players. The main contribution of this paper includes three aspects. First, this paper clarifies that the real vector in a player's self-dependent sufficient statistics is the player's regret on the other player's type. Second, explicit linear program formulations are provided to compute the initial condition of the self-dependent sufficient statistics and the security strategies. Third, in discounted cases when approximated security strategies

are provided, a bound on the performance difference between the approximated security strategy and the security strategy is analyzed.

This paper considers the following two-player zero-sum repeated Bayesian games. Each player has his own type that is only known to himself. The one-stage payoff function depends on both players' types and actions. At the beginning of the game, nature chooses each player's type independently according to some publicly known probability, and send the type to the corresponding player. The players then choose their actions simultaneously based on their own types and both players' history actions at every stage. The reward of the maximizer is the sum of the one-stage payoffs all over the stages in a finite stage game, and the sum of discounted one-stage payoffs in a discounted game.

For the finite horizon case, we first provide an explicit linear program to compute players' security strategies based on the idea of realization plan in sequence form. The computed security strategy depends on both players' history actions. As the time horizon gets long, a huge memory is needed to store the security strategy as the history action space is growing exponentially with respect to the time horizon. To save the memory, we further study the self-dependent sufficient statistics in [1], and show that the real vector in the self-dependent sufficient statistics is the player's regret on the other player's type. Besides, explicit linear programs are presented to compute the initial regrets in the game, and the security strategies based on the self-dependent sufficient statistics. Our simulation results demonstrate that the two security strategies developed from different methods achieve the same game value.

For the discounted case, the self-dependent sufficient statistics consists of the belief on a player's own type and the anti-discounted regret on the other player's type. An approximated security strategy that depends on the self-dependent sufficient statistics is provided, and a linear program is given to compute the approximated security strategy. Moreover, the performance difference between the approximated security strategy and the security strategy is studied, and a bound on the performance difference is presented.

The remainder of this paper is organized as follows. Section II presents the main results for finite horizon games. Section III discusses discounted games. Section IV demonstrates the main results on a jamming problem in underwater sensor networks. Finally, section V provides some future work.

## II. T-STAGE REPEATED BAYESIAN GAMES

Let  $\mathbb{R}^n$  indicate the  $n$ -dimensional real space. For a finite set  $K$ ,  $|K|$  denotes its cardinality, and  $\Delta(K)$  indicates the set of probability distributions over  $K$ .  $\mathbf{1}$  and  $\mathbf{0}$  are appropriately dimensional column vectors with all their elements to be 1 and 0, respectively. Let  $v(0), v(1), \dots$  be a sequence of real values. We adopt the convention that  $\sum_{t=1}^0 v(t) = 0$ , and  $\prod_{t=1}^0 v(t) = 1$ . The supreme norm of a function  $f : D \rightarrow \mathbb{R}$  is defined as  $\|f\|_{\sup} = \sup_{x \in D} |f(x)|$ , where  $D$  is a non-empty set.

A two-player zero-sum repeated Bayesian game is specified by the seven-tuple  $(K, L, A, B, M, p_0, q_0)$ , where

- $K$  and  $L$  are non-empty finite sets, called player 1 and 2's type sets, respectively.
- $A$  and  $B$  are non-empty finite sets, called player 1 and 2's action sets, respectively.
- $M : K \times L \times A \times B \rightarrow \mathbb{R}$  is the one-stage payoff function.  $M^{kl}$  indicates the payoff matrix given player 1's type  $k \in K$  and player 2's type  $l \in L$ . The element  $M_{a,b}^{kl}$  of matrix  $M^{kl}$ , also denoted as  $M(k, l, a, b)$ , is the payoff given player 1's type  $k \in K$  and action  $a \in A$ , and player 2's type  $l \in L$  and action  $b \in B$ .
- $p_0 \in \Delta(K)$  and  $q_0 \in \Delta(L)$  are the initial probabilities on  $K$  and  $L$ , respectively. Without loss of generality, we assume  $p_0^k, q_0^l > 0$  for any  $k \in K$  and  $l \in L$ .

A  $T$ -stage repeated Bayesian game is played as follows. Let  $a_t, b_t$  denote player 1 and player 2's actions at stage  $t = 1, \dots, T$ , respectively. At stage 1,  $k$  and  $l$  are chosen independently according to  $p_0$  and  $q_0$ , and communicated to player 1 and 2, respectively. After the types are chosen, at stage  $t = 1, \dots, T$ , each player chooses his action independently, and announces it publicly. The payoff of player 1 at stage  $t$  is  $M(k, l, a_t, b_t)$ . At stage  $t = 1, \dots, T$ , player 1 and 2's history action sequences  $h_t^A$  and  $h_t^B$  are defined as  $h_t^A = (a_1, \dots, a_{t-1})$  and  $h_t^B = (b_1, \dots, b_{t-1})$ , and their history action spaces are defined as  $H_t^A = A^{t-1}$  and  $H_t^B = B^{t-1}$ , respectively. We assume that  $H_1^A, H_1^B, H_0^A, H_0^B = \emptyset$ . With a little abuse of the terminology  $\in$ , we use  $a_s \in h_t^A$  and  $h_s^A \in h_t^A$  to indicate  $a_s$  and  $h_s^A$  are player 1's action and history action sequence at stage  $s$  in the history action sequence  $h_t^A$  for any  $s = 1, \dots, t-1$ . Similarly,  $b_s \in h_t^B$  and  $h_s^B \in h_t^B$  means that  $b_s$  and  $h_s^B$  are player 2's action and history action sequence at stage  $s$  in the history action sequence  $h_t^B$  for any  $s = 1, \dots, t-1$ .

A behavior strategy for player 1 is an element of  $\sigma = (\sigma_t)_{t=1}^T$ , where  $\sigma_t$  is a map from  $K \times H_t^A \times H_t^B$  to  $\Delta(A)$ . Similarly, a behavior strategy for player 2 is an element of  $\tau = (\tau_t)_{t=1}^T$ , where  $\tau_t$  is a map from  $L \times H_t^A \times H_t^B$  to  $\Delta(B)$ . Denote by  $\Sigma$  and  $\mathcal{T}$  the sets of strategies of player 1 and 2, respectively. Denote by  $\sigma_t^{a_t}(\cdot, \cdot, \cdot)$  and  $\tau_t^{b_t}(\cdot, \cdot, \cdot)$  the probabilities of playing  $a_t$  and  $b_t$  at stage  $t$ , respectively.

A quadruple  $(p_0, q_0, \sigma, \tau)$  induces a probability distribution  $P_{p_0, q_0, \sigma, \tau}$  on the set  $\Omega = K \times L \times (A \times B)^T$  of plays.  $E_{p_0, q_0, \sigma, \tau}$  stands for the corresponding expectation. The payoff with initial probabilities  $p_0, q_0$  and strategies  $\sigma, \tau$  of the  $T$ -stage repeated Bayesian game is defined as

$$\gamma_T(p_0, q_0, \sigma, \tau) = E_{p_0, q_0, \sigma, \tau} \left( \sum_{t=1}^T M(k, l, a_t, b_t) \right).$$

The  $T$ -stage game  $\Gamma_T(p_0, q_0)$  is defined as a two-player zero-sum repeated Bayesian game equipped with initial distribution  $p_0$  and  $q_0$ , strategy spaces  $\Sigma$  and  $\mathcal{T}$ , and payoff function  $\gamma_T(p_0, q_0, \sigma, \tau)$ . In game  $\Gamma_T(p_0, q_0)$ , player 1 wants to find a behavior strategy to *maximize* the payoff  $\gamma_T(p_0, \sigma, \tau)$ , while player 2 wants to *minimize* it.

Consider a  $T$ -stage game  $\Gamma_T(p_0, q_0)$ . The security level  $\underline{V}_T(p_0, q_0)$  of player 1 is defined as  $\underline{V}_T(p_0, q_0) = \max_{\sigma \in \Sigma} \min_{\tau \in \mathcal{T}} \gamma_T(p_0, q_0, \sigma, \tau)$  and the strategy  $\sigma^* \in \Sigma$  which achieves player 1's security level is called the security strategy of player 1. Similarly, the security level  $\bar{V}_T(p_0, q_0)$  of player 2 is defined as  $\bar{V}_T(p_0, q_0) = \min_{\tau \in \mathcal{T}} \max_{\sigma \in \Sigma} \gamma_T(p_0, q_0, \sigma, \tau)$ , and the strategy  $\tau^* \in \mathcal{T}$  which achieves player 2's security level is called the security strategy of player 2. When  $\underline{V}_T(p_0, q_0) = \bar{V}_T(p_0, q_0)$ , we say game  $\Gamma_T(p_0, q_0)$  has a value, i.e. the game has a Nash equilibrium. Since game  $\Gamma_T(p_0, q_0)$  is a finite game, it always has a value denoted by  $V_T(p_0, q_0)$  [9].

#### A. LP formulations for players' security strategies

A  $T$ -stage Bayesian repeated game is a finite game, and its security strategy can be computed by solving a linear program based on the sequence form [11]. The linear program provided in [11], however, can not be directly used, because in our case, the strategies of both players depend on their own types which is not the same situation as in [11]. Therefore, we adopt the idea of *realization plan* in the sequence form, and construct an explicit linear program for  $T$ -stage Bayesian repeated games.

Let's first introduce the *realization plan*. A player's *realization plan* at stage  $t$  given  $(k, l, h_{t+1}^A, h_{t+1}^B)$  is the product of the player's strategy along the path  $(a_1, b_1, a_2, b_2, \dots, a_t, b_t)$ . Define player 1 and 2's realization plan  $x_{k, h_t^A, h_t^B}^{a_t}$  and  $y_{l, h_t^A, h_t^B}^{b_t}$  for  $t = 0, \dots, T$  as

$$x_{k, h_t^A, h_t^B}^{a_t} = p^k \prod_{s=1}^t \sigma_s^{a_s}(k, h_s^A, h_s^B), \quad (1)$$

$$y_{l, h_t^A, h_t^B}^{b_t} = q^l \prod_{s=1}^t \tau_s^{b_s}(l, h_s^A, h_s^B), \quad (2)$$

where  $a_s, h_s^A \in h_t^A$  and  $b_s, h_s^B \in h_t^B$  for all  $s = 1, \dots, t-1$ . It is easy to verify that the joint probability  $P(k, l, h_{t+1}^A, h_{t+1}^B)$  satisfies  $P(k, l, h_{t+1}^A, h_{t+1}^B) = x_{k, h_t^A, h_t^B}^{a_t} y_{l, h_t^A, h_t^B}^{b_t}$ , where  $a_t, h_t^A \in h_{t+1}^A$  and  $b_t, h_t^B \in h_{t+1}^B$ . Let  $x_t = (x_{k, h_t^A, h_t^B})_{k \in K, h_t^A \in H_t^A, h_t^B \in H_t^B}$  and  $y_t = (y_{l, h_t^A, h_t^B})_{l \in L, h_t^A \in H_t^A, h_t^B \in H_t^B}$ . Denote by  $x = (x_t)_{t=1}^T$  and  $y = (y_t)_{t=1}^T$  player 1 and 2's realization plans all over the  $T$ -stage Bayesian game. Player 1's realization plan  $x$  satisfies constraint (3-4), and the corresponding set is denoted by  $X$ . Similarly, player 2's realization plan  $y$  satisfies constraint (5-6), and the corresponding set is denoted by  $Y$ .

$$\mathbf{1}^T x_{k, h_t^A, h_t^B} = x_{k, h_{t-1}^A, h_{t-1}^B}^{a_{t-1}}, \quad \forall t = 1, \dots, T, k \in K, h_t^A \in H_t^A, h_t^B \in H_t^B, \quad (3)$$

$$x_{k, h_t^A, h_t^B} \geq 0, \quad \forall t = 1, \dots, T, k \in K, h_t^A \in H_t^A, h_t^B \in H_t^B, \quad (4)$$

$$\mathbf{1}^T y_{l, h_t^A, h_t^B} = y_{l, h_{t-1}^A, h_{t-1}^B}^{b_{t-1}}, \quad \forall t = 1, \dots, T, l \in L, h_t^A \in H_t^A, h_t^B \in H_t^B, \quad (5)$$

$$y_{l, h_t^A, h_t^B} \geq 0, \quad \forall t = 1, \dots, T, l \in L, h_t^A \in H_t^A, h_t^B \in H_t^B. \quad (6)$$

With perfect recall, for either player, looking for a security strategy is the same as looking for a realization plan that achieves the security level of the player [11].

Given the realization plan of player 1, we define player 1's weighted future security payoff  $u_{l, h_t^A, h_t^B}^{a_t, b_t}(x)$  for  $t = 0, \dots, T$  as

$$u_{l, h_t^A, h_t^B}^{a_t, b_t}(x) = \min_{\tau_{t+1:T}(l) \in \mathcal{T}_{t+1:T}(l)} \sum_{k \in K} x_{k, h_t^A, h_t^B}^{a_t} E \left( \sum_{s=t+1}^T M(k, l, a_s, b_s) | k, l, h_{t+1}^A, h_{t+1}^B \right), \quad (7)$$

where  $h_{t+1}^A = (h_t^A, a_t)$ ,  $h_{t+1}^B = (h_t^B, b_t)$ ,  $\tau_{t+1:T}(l) = (\tau_s(l, :, :))_{s=t+1}^T$ , and  $\mathcal{T}_{t+1:T}(l)$  is the set of player 2's behavior strategies from stage  $t+1$  to  $T$  given player 2's type  $l \in L$ . The pairs  $(h_t^A, a_t)$  and  $(h_t^B, b_t)$  indicate concatenation. Similarly, define player 2's weighted future security payoff  $w_{k, h_t^A, h_t^B}^{a_t, b_t}(y)$  for  $t = 0, \dots, T$  as

$$w_{k, h_t^A, h_t^B}^{a_t, b_t}(y) = \max_{\sigma_{t+1:T}(k) \in \Sigma_{t+1:T}(k)} \sum_{l \in L} y_{l, h_t^A, h_t^B}^{b_t} E \left( \sum_{s=t+1}^T M(k, l, a_s, b_s) | k, l, h_{t+1}^A, h_{t+1}^B \right), \quad (8)$$

where  $\sigma_{t+1:T}(k) = (\sigma_s(k, :, :))_{s=t+1}^T$ , and  $\Sigma_{t+1:T}(k)$  is the set of player 1's strategies from stage  $t+1$  to  $T$  given player 1's type  $k \in K$ .

For  $t = 1, \dots, T$ ,  $u_{l, h_t^A, h_t^B}(x)$ ,  $w_{k, h_t^A, h_t^B}(y)$  are  $|A| \times |B|$  matrices whose elements are  $u_{l, h_t^A, h_t^B}^{a_t, b_t}(x)$  and  $w_{k, h_t^A, h_t^B}^{a_t, b_t}(y)$ , respectively. For  $t = 0$ , since  $a_t, b_t, h_t^A, h_t^B \in \emptyset$ ,  $u_{l, h_t^A, h_t^B}(x)$  and  $w_{k, h_t^A, h_t^B}(y)$  are scalars, and denoted as  $u_{l,0}(x)$  and  $w_{k,0}(y)$ , respectively. Define  $u_t(x) = (u_{l, h_t^A, h_t^B}(x))_{l \in L, h_t^A \in H_t^A, h_t^B \in H_t^B}$ , and  $u(x) = (u_t(x))_{t=1}^{T-1}$ . Similarly, define  $w_t(y) = (w_{k, h_t^A, h_t^B}(y))_{k \in K, h_t^A \in H_t^A, h_t^B \in H_t^B}$ , and  $w(y) = (w_t(y))_{t=1}^{T-1}$ . For the convenience of the rest of this paper, let  $U$  and  $W$  be

the real spaces of appropriate dimensions which  $u$  and  $w$  take values in. The weighted future security payoffs  $u, w$  satisfy backward recursive formulas.

**Lemma 1.** Consider a  $T$ -stage Bayesian game  $\Gamma_T(p, q)$ . Player 1 and 2's weighted future security payoffs  $u_{l, h_t^A, h_t^B}^{a_t, b_t}(x)$  and  $w_{k, h_t^A, h_t^B}^{a_t, b_t}(y)$  defined in (7) and (8) satisfy

$$u_{l, h_t^A, h_t^B}^{a_t, b_t}(x) = \min_{\tau_{t+1}(l, h_{t+1}^A, h_{t+1}^B) \in \Delta(B)} \left( \sum_{k \in K} x_{k, h_{t+1}^A, h_{t+1}^B}^T M^{kl} + \mathbf{1}^T u_{l, h_{t+1}^A, h_{t+1}^B}(x) \right) \tau_{t+1}(l, h_{t+1}^A, h_{t+1}^B), \quad (9)$$

$$w_{k, h_t^A, h_t^B}^{a_t, b_t}(y) = \max_{\sigma_{t+1}(k, h_{t+1}^A, h_{t+1}^B) \in \Delta(A)} \sigma_{t+1}(k, h_{t+1}^A, h_{t+1}^B)^T \left( \sum_{l \in L} M^{kl} y_{l, h_{t+1}^A, h_{t+1}^B} + w_{k, h_{t+1}^A, h_{t+1}^B}(y) \mathbf{1} \right), \quad (10)$$

for all  $t = 0, \dots, T-1$ , where  $h_{t+1}^A = (h_t^A, a_t)$ ,  $h_{t+1}^B = (h_t^B, b_t)$ ,  $x_{k, h_{t+1}^A, h_{t+1}^B} \in \mathbb{R}^{|A|}$  is player 1's realization plan whose element is defined as in (1), and  $y_{l, h_{t+1}^A, h_{t+1}^B} \in \mathbb{R}^{|B|}$  is player 2's realization plan whose element is defined as in (2). Here,  $(h_t^A, a_t)$  and  $(h_t^B, b_t)$  indicate concatenation.

*Proof.* According to equation (7), we have

$$\begin{aligned} u_{l, h_{T-1}^A, h_{T-1}^B}^{a_{T-1}, b_{T-1}}(x) &= \min_{\tau_T(l, h_T^A, h_T^B) \in \Delta(B)} \sum_{k \in K} x_{k, h_T^A, h_T^B}^{a_{T-1}, b_{T-1}} \sigma_T^T(k, h_T^A, h_T^B) M^{kl} \tau_T(l, h_T^A, h_T^B) \\ &= \min_{\tau_T(l, h_T^A, h_T^B) \in \Delta(B)} \sum_{k \in K} x_{k, h_T^A, h_T^B}^T M^{kl} \tau_T(l, h_T^A, h_T^B) \\ &= \min_{\tau_T(l, h_T^A, h_T^B) \in \Delta(B)} \left( \sum_{k \in K} x_{k, h_T^A, h_T^B}^T M^{kl} + \mathbf{1}^T u_{l, h_T^A, h_T^B}(x) \right) \tau_T(l, h_T^A, h_T^B), \end{aligned}$$

where the last equality holds because  $u_{l, h_T^A, h_T^B}(x)$  is a zero matrix according to (7).

Suppose equation (9) holds for all  $t = 1, \dots, T-1$ . Consider the case of  $t-1$ .

$$\begin{aligned} u_{l, h_{t-1}^A, h_{t-1}^B}^{a_{t-1}, b_{t-1}}(x) &= \min_{\tau_{t:T}(l) \in \mathcal{T}_{t:T}(l)} \sum_{k \in K} x_{k, h_{t-1}^A, h_{t-1}^B}^{a_{t-1}, b_{t-1}} \left( \sigma_t^T(k, h_t^A, h_t^B) M^{kl} \tau_t(l, h_t^A, h_t^B) \right. \\ &\quad \left. + \sum_{a_t \in A} \sum_{b_t \in B} P(a_t, b_t | k, l, h_t^A, h_t^B) E \left( \sum_{s=t+1}^T M(k, l, a_s, b_s) | k, l, h_t^A, h_t^B, a_t, b_t \right) \right) \\ &= \min_{\tau_t(l, h_t^A, h_t^B) \in \Delta(B)} \left\{ \sum_{k \in K} x_{k, h_t^A, h_t^B}^T M^{kl} \tau_t(l, h_t^A, h_t^B) \right. \\ &\quad \left. + \sum_{a_t \in A} \sum_{b_t \in B} \tau_t^{b_t}(l, h_t^A, h_t^B) \min_{\tau_{t+1:T}(l) \in \mathcal{T}_{t+1:T}(l)} \sum_{k \in K} x_{k, h_t^A, h_t^B}^{a_t} E \left( \sum_{s=t+1}^T M(k, l, a_s, b_s) | k, l, h_{t+1}^A, h_{t+1}^B \right) \right\} \\ &= \min_{\tau_t(l, h_t^A, h_t^B) \in \Delta(B)} \left( \sum_{k \in K} x_{k, h_t^A, h_t^B}^T M^{kl} + \mathbf{1}^T u_{l, h_t^A, h_t^B}(x) \right) \tau_t(l, h_t^A, h_t^B). \end{aligned}$$

Therefore, equation (9) holds for all  $t = 0, \dots, T-1$ .

Following the same steps, we show that equation (10) is also true.  $\square$

Now, we are ready to present the explicit LP formulations for both players.

**Theorem 2.** Consider a  $T$ -stage repeated Bayesian game  $\Gamma_T(p, q)$ . The game value  $V_T(p, q)$  satisfies

$$V_T(p, q) = \max_{x \in X, u \in U, u_{:,0} \in \mathbb{R}^{|L|}} \sum_{l \in L} q^l u_{l,0} \quad (11)$$

$$\text{s.t. } \sum_{k \in K} M^{klT} x_{k, h_1^A, h_1^B} + u_{l, h_1^A, h_1^B}^T \mathbf{1} \geq u_{l,0} \mathbf{1}, \quad \forall l \in L, \quad (12)$$

$$\sum_{k \in K} M^{klT} x_{k, h_{t+1}^A, h_{t+1}^B} + u_{l, h_{t+1}^A, h_{t+1}^B}^T \mathbf{1} \geq u_{l, h_t^A, h_t^B}^{a_t, b_t} \mathbf{1}, \quad \forall t = 1, \dots, T-1, l \in L, h_t^A \in H_t^A, h_t^B \in H_t^B, \quad (13)$$

where  $u_{l,h_T^A,h_T^B}$  is a zero matrix for all  $l \in L$ ,  $X$  is a set including all real vectors satisfying (3-4), and  $U$  is a real space of appropriate dimension. Player 1's security strategy  $\sigma_t^{a_t^*}(k, h_t^A, h_t^B)$  for all  $t = 1, \dots, T$ ,  $k \in K$ ,  $h_t^A \in H_t^A$ ,  $h_t^B \in H_t^B$ , and  $a_t \in A$  satisfies

$$\sigma_t^{a_t^*}(k, h_t^A, h_t^B) = \frac{x_{k,h_t^A,h_t^B}^{a_t^*}}{x_{k,h_{t-1}^A,h_{t-1}^B}^{a_{t-1}^*}}. \quad (14)$$

Dually, the game value  $V_T(p, q)$  also satisfies

$$V_T(p, q) = \min_{y \in Y, w \in W, w_{:,0} \in \mathbb{R}^{|K|}} \sum_{k \in K} p^k w_{k,0} \quad (15)$$

$$s.t. \sum_{l \in L} M^{kl} y_{l,h_1^A,h_1^B} + w_{k,h_1^A,h_1^B} \mathbf{1} \leq w_{k,0} \mathbf{1}, \quad \forall k \in K, \quad (16)$$

$$\sum_{l \in L} M^{kl} y_{l,h_{t+1}^A,h_{t+1}^B} + w_{k,h_{t+1}^A,h_{t+1}^B} \mathbf{1} \leq w_{k,h_t^A,h_t^B}^{a_t,b_t} \mathbf{1}, \quad \forall t = 1, \dots, T-1, k \in K, h_t^A \in H_t^A, h_t^B \in H_t^B, \quad (17)$$

where  $w_{k,h_T^A,h_T^B}$  is a zero matrix for all  $k \in K$ ,  $Y$  is a set including all real vectors satisfying (5-6), and  $W$  is a real space of appropriate dimension. Player 2's security strategy  $\tau_t^{b_t^*}(l, h_t^A, h_t^B)$  for all  $t = 1, \dots, T$ ,  $l \in L$ ,  $h_t^A \in H_t^A$ ,  $h_t^B \in H_t^B$ , and  $a_t \in A$  satisfies

$$\tau_t^{b_t^*}(l, h_t^A, h_t^B) = \frac{y_{l,h_t^A,h_t^B}^{b_t^*}}{y_{l,h_{t-1}^A,h_{t-1}^B}^{b_{t-1}^*}}. \quad (18)$$

*Proof.* Equation (7) indicates that  $V_T(p, q) = \max_{x \in X} \sum_{l \in L} q^l u_{l,0}(x)$ , where  $u_{l,0}(x)$  satisfies (9).

According to the duality theory in LP problem, equation (9) can be rewritten as

$$u_{l,h_t^A,h_t^B}^{a_t,b_t}(x) = \max_{u_{l,h_t^A,h_t^B}^{a_t,b_t} \in \mathbb{R}} u_{l,h_t^A,h_t^B}^{a_t,b_t} \quad (19)$$

$$s.t. \sum_{k \in K} M^{klT} x_{k,h_{t+1}^A,h_{t+1}^B} + u_{l,h_{t+1}^A,h_{t+1}^B}^T(x) \mathbf{1} \geq u_{l,h_t^A,h_t^B}^{a_t,b_t} \mathbf{1}, \quad (20)$$

for  $t = 0, \dots, T-1$ .

For  $t = T-1$ , since  $u_{l,h_T^A,h_T^B}(x)$  is a zero matrix, we have

$$u_{l,h_{T-1}^A,h_{T-1}^B}^{a_{T-1},b_{T-1}}(x) = \max_{u_{l,h_{T-1}^A,h_{T-1}^B}^{a_{T-1},b_{T-1}} \in \mathbb{R}} u_{l,h_{T-1}^A,h_{T-1}^B}^{a_{T-1},b_{T-1}}, \quad (21)$$

$$s.t. \sum_{k \in K} M^{klT} x_{k,h_T^A,h_T^B} \geq u_{l,h_{T-1}^A,h_{T-1}^B}^{a_{T-1},b_{T-1}} \mathbf{1}. \quad (22)$$

For  $t = T-2$ , we define

$$\hat{u}_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}}(x) = \max_{u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}} \in \mathbb{R}, u_{l,(h_{T-2}^A,a_{T-2}),(h_{T-2}^B,b_{T-2})} \in \mathbb{R}^{|A| \times |B|}} u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}} \quad (23)$$

$$s.t. \sum_{k \in K} M^{klT} x_{k,h_{T-1}^A,h_{T-1}^B} + u_{l,(h_{T-2}^A,a_{T-2}),(h_{T-2}^B,b_{T-2})}^T \mathbf{1} \geq u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}} \mathbf{1}, \quad (24)$$

$$\sum_{k \in K} M^{klT} x_{k,h_T^A,h_T^B} \geq u_{l,(h_{T-2}^A,a_{T-2}),(h_{T-2}^B,b_{T-2})}^{a_{T-1},b_{T-1}} \mathbf{1}, \forall a_{T-1} \in A, b_{T-1} \in B. \quad (25)$$

We will show that  $u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}}(x) = \hat{u}_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}}(x)$ . Equation (19) implies that

$$u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}}(x) = \max_{u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}} \in \mathbb{R}} u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}} \quad (26)$$

$$s.t. \sum_{k \in K} M^{klT} x_{k,h_{T-1}^A,h_{T-1}^B} + u_{l,(h_{T-2}^A,a_{T-2}),(h_{T-2}^B,b_{T-2})}^{*T} \mathbf{1} \geq u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}} \mathbf{1}, \quad (27)$$

$$(28)$$

where the element in  $u_{l,(h_{T-2}^A,a_{T-2}),(h_{T-2}^B,b_{T-2})}^{*T}$  is the corresponding maximum of LP (21-22).

It is easy to see that the feasible area of (23-25) is included in the nested LP (26-27), and hence  $u_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}}(x) \geq \hat{u}_{l,h_{T-2}^A,h_{T-2}^B}^{a_{T-2},b_{T-2}}(x)$ .

Meanwhile, let  $u_{l, h_{T-2}^A, h_{T-2}^B}^{a_{T-2}, b_{T-2}, *}$ ,  $u_{l, (h_{T-2}^A, a_{T-2}), (h_{T-2}^B, b_{T-2})}^*$  be the optimal solution to the nested LP (26-27). It is easy to check that  $u_{l, h_{T-2}^A, h_{T-2}^B}^{a_{T-2}, b_{T-2}, *}$ ,  $u_{l, (h_{T-2}^A, a_{T-2}), (h_{T-2}^B, b_{T-2})}^*$  is a feasible solution to LP (23-25), and hence  $u_{l, h_{T-2}^A, h_{T-2}^B}^{a_{T-2}, b_{T-2}}(x) \leq \hat{u}_{l, h_{T-2}^A, h_{T-2}^B}^{a_{T-2}, b_{T-2}}(x)$ . Therefore, we have  $u_{l, h_{T-2}^A, h_{T-2}^B}^{a_{T-2}, b_{T-2}}(x) = \hat{u}_{l, h_{T-2}^A, h_{T-2}^B}^{a_{T-2}, b_{T-2}}(x)$ .

Following the same steps, we can show the case for  $t = T - 3, \dots, 0$ . Therefore, we have

$$u_{l,0}(x) = \max_{u_{l,0} \in \mathbb{R}, u \in U} u_{l,0} \quad (29)$$

$$s.t. \sum_{k \in K} M^{klT} x_{k, h_1^A, h_1^B} + u_{l, h_1^A, h_1^B}^T \mathbf{1} \geq u_{l,0} \mathbf{1}, \quad (30)$$

$$\sum_{k \in K} M^{klT} x_{k, h_{t+1}^A, h_{t+1}^B} + u_{l, h_{t+1}^A, h_{t+1}^B}^T \mathbf{1} \geq u_{l, h_t^A, h_t^B}^{a_t, b_t} \mathbf{1}, \forall t = 1, \dots, T-1, h_t^A \in H_t^A, h_t^B \in H_t^B, \quad (31)$$

and equation (11-13) is shown. From the definition of  $x$  in (1), we derive player 1's security strategy as in (14).

Following the same steps, we show that equation (15-17) is also true, and player 2's security strategy is computed as in (18).  $\square$

The sizes of the LP formulations in (11-13) and (15-17) are both linear in the size of the game tree, i.e. linear in the sizes of both players' type sets, polynomial in the sizes of both players' action sets, and exponential in the time horizon. Let us first analyze the variable size in the linear program (11-13). Variable  $x = (x_t)_{t=1}^T$  is defined for every stage, where  $x_t = (x_{k, h_t^A, h_t^B})_{k \in K, h_t^A \in H_t^A, h_t^B \in H_t^B}$  is defined for any  $k \in K$ ,  $h_t^A \in H_t^A$ , and  $h_t^B \in H_t^B$ . Together with the fact that  $x_{k, h_t^A, h_t^B} \in \mathbb{R}^{|A|}$ , we see that variable  $x$  consists of  $|A||K|(1 + |A||B| + \dots + (|A||B|)^{T-1}) = O(|K||A|^{T+1}|B|^T)$  scalar variables. Variable  $u = (u_t)_{t=1}^{T-1}$ , where  $u_t = (u_{l, h_t^A, h_t^B})_{l \in L, h_t^A \in H_t^A, h_t^B \in H_t^B}$  whose element is a  $|A| \times |B|$  matrix, is defined for any  $l \in L$ ,  $h_t^A \in H_t^A$  and  $h_t^B \in H_t^B$  at stage  $t = 1, \dots, T-1$ , and hence consists of  $|A||B||L|(1 + |A||B| + \dots + (|A||B|)^{T-2}) = O(|L||A|^T|B|^T)$  scalar variables. Variable  $u_{:,0} \in \mathbb{R}^{|L|}$  consists of  $|L|$  scalar variables. In all, there are  $O((|K| + |L|)(|A||B|)^{T+1})$  scalar variables. Next, we take a look at the constraints in LP formulation (11-13). Constraint (12) has  $|B||L|$  inequalities. Constraint (13) has  $|B||L|(|A||B| + \dots + (|A||B|)^{T-1}) = O(|L||A|^T|B|^{T+1})$  inequalities. Constraint (3) has  $|K|(1 + \dots + (|A||B|)^{T-1}) = O(|K||A|^T|B|^T)$  inequalities. Constraint (4) has  $|A||K|(1 + \dots + (|A||B|)^{T-1}) = O(|K||A|^{T+1}|B|^T)$  inequalities. In all, there are  $O((|K| + |L|)(|A||B|)^{T+1})$  inequalities. Therefore, we say that the LP formulation (11-13) has its size linear in  $|K|$  and  $|L|$ , polynomial in  $|A|$  and  $|B|$ , and exponential in  $T$ . Since the LP formulation (15-17) is dual to the LP formulation (11-13), the two LP formulations have the same sizes.

## B. Security strategies based on fixed-sized sufficient statistics and dual games

Theorem 2 provides LP formulations to compute both players' security strategies which depend on both players' history actions. Notice that history action space grows exponentially on time horizon. As time horizon gets long, players need a great amount of memories to record players' history action space and the corresponding security strategy. This subsection will examine players' security strategies that depend on fixed-sized sufficient statistics to save memories.

It was shown that a player's security strategy in the dual game with some special initial parameters is also the player's security strategy in the primal game, and the security strategy only depends on a fixed-sized sufficient statistics [12], [9]. This subsection clarifies what the special initial parameters in dual games mean in the primal game, and give LP formulation and algorithms to compute the initial parameters and the corresponding security strategies.

First of all, we would like to introduce *two dual games* of a  $T$ -stage repeated Bayesian game  $\Gamma_T(p, q)$ . Game  $\Gamma_T(p, q)$ 's type 1 dual game  $\tilde{\Gamma}_T^1(\mu, q)$  is defined with respect to its first parameter  $p$ , where  $\mu \in \mathbb{R}^{|K|}$  is called the initial regret with respect to player 1's type. The dual game  $\tilde{\Gamma}_T^1(\mu, q)$  is played as follows. Player 1 chooses  $k$  without informing player 2. Independently, nature chooses player 2's type according to  $q$ , and announces it to player 2 only. From stage 1 to  $T$ , knowing both players' history actions, both players choose actions simultaneously. Let  $p$  be player 1's strategy to choose his own type, and  $\sigma \in \Sigma$  and  $\tau \in \mathcal{T}$  be player 1 and 2's strategies to choose actions. Player 1's payoff  $\tilde{\gamma}_T^1(\mu, q, p, \sigma, \tau)$  is defined as

$$\tilde{\gamma}_T^1(\mu, q, p, \sigma, \tau) = E_{p, q, \sigma, \tau} \left( \mu^k + \sum_{t=1}^T M(k, l, a_t, b_t) \right). \quad (32)$$

We can see that the main difference between the type 1 dual game and the primal game is that in type 1 dual game, player 1 has an initial regret instead of a initial probability  $p$ , and he himself instead of the nature chooses his own type.

Similarly, the type 2 dual game  $\tilde{\Gamma}_T^2(p, \nu)$  is defined with respect to the second parameter  $q$ , where  $\nu \in \mathbb{R}^{|L|}$  is called the initial regret with respect to player 2's type. The dual game  $\tilde{\Gamma}_T^2(p, \nu)$  is played as follows. Player 2 chooses  $l$  without informing player 1. Meanwhile, player 1's type is chosen according to  $p$ , and is only announced to player 1. From stage 1 to  $T$ , knowing

both players' history actions, both players choose actions independently. Let  $q$  be player 2's strategy to choose his type  $l$ . Player 1's payoff  $\tilde{\gamma}_T^2(p, \nu, q, \sigma, \tau)$  is defined as

$$\tilde{\gamma}_T^2(p, \nu, q, \sigma, \tau) = E_{p,q,\sigma,\tau} \left( \nu^l + \sum_{t=1}^T M(k, l, a_t, b_t) \right). \quad (33)$$

In both dual games, player 1 wants to maximize the payoff, while player 2 wants to minimize it.

Both dual games are finite, and hence have game values denoted by  $\tilde{V}_T^1(\mu, q)$  and  $\tilde{V}_T^2(p, \nu)$ . They are related to the game value of the primal game in the following way [9].

$$\tilde{V}_T^1(\mu, q) = \max_{p \in \Delta(K)} \{V_T(p, q) + p^T \mu\}, \quad (34)$$

$$V_T(p, q) = \min_{\mu \in \mathbb{R}^{|K|}} \{\tilde{V}_T^1(\mu, q) - p^T \mu\}, \quad (35)$$

$$\tilde{V}_T^2(p, \nu) = \min_{q \in \Delta(L)} \{V_T(p, q) + q^T \nu\}, \quad (36)$$

$$V_T(p, q) = \max_{\nu \in \mathbb{R}^{|L|}} \{\tilde{V}_T^2(q, \nu) - q^T \nu\}. \quad (37)$$

Let  $\mu^*$  and  $\nu^*$  be the solutions to the optimal problems on the right hand side of (35) and (37), respectively. Player 2's security strategy in the type 1 dual game  $\tilde{\Gamma}_T^1(\mu^*, q)$  is his security strategy in the primal game  $\Gamma_T(p, q)$ , and player 1's security strategy in the type 2 dual game  $\tilde{\Gamma}_T^2(p, \nu^*)$  is also his security strategy in the primal game  $\Gamma_T(p, q)$  [1].

The next questions are what  $\mu^*$  and  $\nu^*$  are, and how to compute them. To answer these questions, we have the following lemma.

**Lemma 3.** Consider a  $T$ -stage repeated Bayesian game  $\Gamma_T(p, q)$ . Let  $\sigma_{p,q}^*$  and  $\tau_{p,q}^*$  be player 1 and 2's security strategies in  $\Gamma_T(p, q)$ , respectively. Denote by  $x_{p,q}^*$  and  $y_{p,q}^*$  the corresponding optimal realization plans of player 1 and 2. The optimal solution  $\mu^*$  to the optimal problem  $\min_{\mu \in \mathbb{R}^{|K|}} \{\tilde{V}_T^1(\mu, q) - p^T \mu\}$  is

$$\mu^{*k} = -w_{k,0}(y_{p,q}^*), \forall k \in K \quad (38)$$

where  $w_{k,0}(y_{p,q}^*) = w_{k,h_0^A,h_0^B}^{a_0,b_0}(y_{p,q}^*)$ , which is defined in (8) and computed according to the linear program (15-17).

The optimal solution  $\nu^*$  to the optimal problem  $\max_{\nu \in \mathbb{R}^{|L|}} \{\tilde{V}_T^2(q, \nu) - q^T \nu\}$  is

$$\nu^{*l} = -u_{l,0}(x_{p,q}^*), \forall l \in L \quad (39)$$

where  $u_{l,0}(x_{p,q}^*) = u_{l,h_0^A,h_0^B}^{a_0,b_0}(x_{p,q}^*)$ , which is defined in (7) and computed according to the linear program (11-13).

*Proof.* It is easy to verify that

$$V_T(p, q) = p^T w_{:,0}(y_{p,q}^*) = -p^T \mu^*. \quad (40)$$

Next, we show that

$$\tilde{V}_T^1(\mu^*, q) = 0. \quad (41)$$

Equation (34) implies that  $\tilde{V}_T^1(\mu^*, q) = \max_{p' \in \Delta(K)} \{V_T(p', q) + p'^T \mu^*\} \geq V_T(p, q) + p^T \mu^* = 0$ .

Meanwhile, for any  $p' \in \Delta(K)$ , we have

$$\begin{aligned} V_T(p', q) &= \min_{y \in Y} \max_{\sigma \in \Sigma} E_{p',q,\sigma,y} \left( \sum_{t=1}^T M(k, l, a_t, b_t) \right) \leq \max_{\sigma \in \Sigma} E_{p',q,\sigma,y_{p,q}^*} \left( \sum_{t=1}^T M(k, l, a_t, b_t) \right) \\ &= \sum_{k \in K} p'^k \max_{\sigma(k) \in \Sigma(k)} E_{q,\sigma(k),y_{p,q}^*} \left( \sum_{t=1}^T M(k, l, a_t, b_t) | k \right) \\ &= \sum_{k \in K} p'^k w_{k,0}(y_{p,q}^*) = -p'^T \mu^*, \end{aligned}$$

which implies that  $V_T(p', q) + p'^T \mu^* \leq 0$  for any  $p' \in \Delta(K)$ . Hence,  $\tilde{V}_T^1(\mu^*, q) \leq 0$  according to (34). Therefore, equation (41) is true.

Equation (41) implies that  $\tilde{V}_T^1(\mu^*, q) - p^T \mu^* = -p^T \mu^* = V_T(p, q)$ , where the second equality is based on (40). According to equation (35), we see that  $-w_{k,0}(y_{p,q}^*)$  is an optimal solution to the optimal problem on the right hand side of (35). From the proof of Theorem 2, we see that  $w_{:,0}(y_{p,q}^*) = w_{:,0}^*$ , where  $w_{:,0}^*$  is the optimal solution to the linear program (15-17).

Following the same steps, we show that

$$\tilde{V}_T^2(q, \nu^*) = 0, \quad (42)$$

and  $-u_{l,0}(x_{p,q}^*)$  is an optimal solution to the optimal problem on the right hand side of (37). Moreover,  $u_{:,0}(x_{p,q}^*)$  equals to  $u_{:,0}^*$  the optimal solution to the linear program (11-13), according to the proof of Theorem 2.  $\square$

The parameter  $\mu^{*k}$  can be seen as player 2' initial regret given player 1's type  $k$ , and  $\nu^{*l}$  is player 1's initial regret given player 2's type  $l$ . Now that we have figured out the two special parameters in the dual game, our next step is to study *the security strategies in the dual games*. In type 1 dual game  $\tilde{\Gamma}_T^1(\mu, q)$ , player 2 keeps track of two variables, the *belief state*  $q_t \in \Delta(L)$  on his own type, and his *regret*  $\mu_t \in \mathbb{R}^{|K|}$  on player 1's type. The *belief on player 2's type* is defined as

$$q_t^l = P(l|k, h_t^A, h_t^B), \forall l \in L, t = 1, \dots, T, \quad (43)$$

and is updated as follows

$$q_{t+1}^l = q^{+l}(b_t, z_t, q_t) = \frac{q_t^l z_t^l(b_t)}{\bar{z}_{q_t, z_t}(b_t)}, \forall l \in L, \text{ with } q_1 = q, \quad (44)$$

where  $z_t^l = \tau_t(l, h_t^A, h_t^B) \in \Delta(B)$ , and  $\bar{z}_{q_t, z_t}(b_t) = \sum_{l \in L} q_t^l z_t^l(b_t)$ . The *regret on player 1's type* is defined as

$$\mu_t^k = \mu^k + \sum_{s=1}^{t-1} E(M(k, l, a_s, b_s) | k, h_{s+1}^A, h_{s+1}^B), \forall k \in K, t = 1, \dots, T, \quad (45)$$

and is updated as follows

$$\mu_{t+1}^k = \mu^{+k}(\mu_t, a_t, b_t, z_t, q_t) = \mu_t^k + E(M(k, l, a_t, b_t) | k, h_{t+1}^A, h_{t+1}^B) = \mu_t^k + \sum_{l \in L} q_{t+1}^l M_{a_t, b_t}^{kl}, \forall k \in K, \text{ with } \mu_1^k = \mu^k. \quad (46)$$

Player 2's security strategy at stage  $t$  in  $\tilde{\Gamma}_T^1(\mu, q)$  can be computed based on the backward recursive equation (47), and depends only on  $t$ ,  $\mu_t$  and  $q_t$  [9].

$$\tilde{V}_n^1(\mu_t, q_t) = \min_{z \in \Delta(B)^{|L|}} \max_{a \in A} \sum_{b \in B} \bar{z}_{q_t, z}(b) \tilde{V}_{n-1}^1(\mu^+( \mu_t, a, b, z, q_t), q^+(b, z, q_t)), \quad (47)$$

where  $n = T + 1 - t$ .

Similarly, in type 2 dual game  $\tilde{\Gamma}_T^2(p, \nu)$ , player 1 also records two variables, the *belief*  $p_t \in \Delta(K)$  on player 1's type and the *regret*  $\nu_t \in \mathbb{R}^{|L|}$  on player 2's type. The *belief on player 1's type* is defined as

$$p_t^k = P(k|l, h_t^A, h_t^B), \forall k \in K, t = 1, \dots, T, \quad (48)$$

and is updated as below

$$p_{t+1}^k = p^{+k}(a_t, r_t, p_t) = \frac{p_t^k r_t^k(a_t)}{\bar{r}_{p_t, r_t}(a_t)}, \forall k \in K, \text{ with } p_1 = p, \quad (49)$$

where  $r_t^k = \sigma_t(k, h_t^A, h_t^B) \in \Delta(A)$ , and  $\bar{r}_{p_t, r_t}(a_t) = \sum_{k \in K} p_t^k r_t^k(a_t)$ . The *regret on player 2's type* is defined as

$$\nu_t^l = \nu^l + \sum_{s=1}^{t-1} E(M(k, l, a_s, b_s) | l, h_{s+1}^A, h_{s+1}^B), \forall l \in L, t = 1, \dots, T, \quad (50)$$

and is updated as below

$$\nu_{t+1}^l = \nu^{+l}(\nu_t, a_t, b_t, r_t, p_t) = \nu_t^l + E(M(k, l, a_t, b_t) | l, h_{t+1}^A, h_{t+1}^B) = \nu_t^l + \sum_{k \in K} p_{t+1}^k M_{a_t, b_t}^{kl}, \forall l \in L, \text{ with } \nu_1^l = \nu^l. \quad (51)$$

Player 1's security strategy at stage  $t$  in  $\tilde{\Gamma}_T^2(p, \nu)$  can be computed based on the backward recursive equation (52), and depends only on  $t$ ,  $p_t$  and  $\nu_t$  [9].

$$\tilde{V}_n^2(p_t, \nu_t) = \max_{r \in \Delta(A)^{|K|}} \min_{b \in B} \sum_{a \in A} \bar{r}_{p_t, r}(a) \tilde{V}_{n-1}^2(p^+(a, r, p_t), \nu^+(\nu_t, a, b, r, p_t)), \quad (52)$$

where  $n = T + 1 - t$ .

From the analysis above, we see that the security strategies of player 1 and 2 in the corresponding dual games depend only on the fixed-sized sufficient statistics,  $(t, p_t, \nu_t)$  and  $(t, \mu_t, q_t)$ , respectively, at stage  $t$ . Moreover, the sufficient statistics  $(t, p_t, \nu_t)$  and  $(t, \mu_t, q_t)$  are fully accessible to player 1 and 2, respectively, in the corresponding dual games. Based on the LP formulation of  $V_T(p, q)$ , we give the LP formulations to compute player 1's security strategy in type 2 dual game  $\tilde{\Gamma}_T^2(p, \nu)$  and player 2's security strategy in type 1 dual game  $\tilde{\Gamma}_T^1(\mu, q)$  as follows.



**Theorem 4.** Consider type 2 dual game  $\tilde{\Gamma}_T^2(p, \nu)$ . Let  $p_t$  and  $\nu_t$  be the belief on player 1's type and the regret on player 2's type at stage  $t$ , respectively. The game value  $\tilde{V}_n^2(p_t, \nu_t)$  of  $n$  stage type 2 dual game  $\tilde{\Gamma}_n^2(p_t, \nu_t)$  satisfies the following LP formulation, where  $n = T + 1 - t$ .

$$\tilde{V}_n^2(p_t, \nu_t) = \max_{x \in X, u \in U, u_{:,0} \in \mathbb{R}^{|L|}, \tilde{u} \in \mathbb{R}} \tilde{u} \quad (53)$$

$$s.t. u_{:,0} + \nu_t \geq \tilde{u} \mathbf{1} \quad (54)$$

$$\sum_{k \in K} M^{klT} x_{k, h_1^A, h_1^B} + u_{l, h_1^A, h_1^B}^T \mathbf{1} \geq u_{l,0} \mathbf{1}, \quad \forall l \in L, \quad (55)$$

$$\sum_{k \in K} M^{klT} x_{k, h_{t+1}^A, h_{t+1}^B} + u_{l, h_{t+1}^A, h_{t+1}^B}^T \mathbf{1} \geq u_{l, h_t^A, h_t^B}^{a_t, b_t} \mathbf{1}, \quad \forall t = 1, \dots, n-1, l \in L, h_t^A \in H_t^A, h_t^B \in H_t^B, \quad (56)$$

where  $u_{l, h_n^A, h_n^B}$  is a zero matrix for all  $l \in L$ ,  $X$  is a set including all real vectors satisfying (3-4) with  $x_{k, h_0^A, h_0^B}^{a_0} = p_t^k$ , and  $U$  is a real space of appropriate dimension. Player 1's security strategy  $\tilde{\sigma}_t^*(k, p_t, \nu_t)$  at stage  $t$  is

$$\tilde{\sigma}_t^*(k, p_t, \nu_t) = \frac{x_{k, h_1^A, h_1^B}^*}{p_t^k}. \quad (57)$$

Similarly, for type 1 dual game  $\tilde{\Gamma}_T^1(\mu, q)$ , let  $\mu_t$  and  $q_t$  be the regret on player 1's type and the belief on player 2's type at stage  $t$ . The game value  $\tilde{V}_n^1(\mu_t, q_t)$  of  $n$  stage type 1 dual game  $\tilde{\Gamma}_n^1(\mu_t, q_t)$  satisfies the following LP formulation, where  $n = T + 1 - t$ .

$$\tilde{V}_n^1(\mu_t, q_t) = \min_{y \in Y, w \in W, w_{:,0} \in \mathbb{R}^{|K|}, \tilde{w} \in \mathbb{R}} \tilde{w} \quad (58)$$

$$s.t. w_{:,0} + \mu_t \leq \tilde{w} \mathbf{1}, \quad (59)$$

$$\sum_{l \in L} M^{kl} y_{l, h_1^A, h_1^B} + w_{k, h_1^A, h_1^B} \mathbf{1} \leq w_{k,0} \mathbf{1}, \quad \forall k \in K, \quad (60)$$

$$\sum_{l \in L} M^{kl} y_{l, h_{t+1}^A, h_{t+1}^B} + w_{k, h_{t+1}^A, h_{t+1}^B} \mathbf{1} \leq w_{k, h_t^A, h_t^B}^{a_t, b_t} \mathbf{1}, \quad \forall t = 1, \dots, n-1, k \in K, h_t^A \in H_t^A, h_t^B \in H_t^B, \quad (61)$$

where  $w_{k, h_n^A, h_n^B}$  is a zero matrix for all  $k \in K$ ,  $Y$  is a set including all real vectors satisfying (5-6) with  $y_{l, h_0^A, h_0^B}^{b_0} = q_t^l$ , and  $W$  is a real space of appropriate dimension. Player 2's security strategy  $\tilde{\tau}_t^*(l, \mu_t, q_t)$  at stage  $t$  is

$$\tilde{\tau}_t^*(l, \mu_t, q_t) = \frac{y_{l, h_1^A, h_1^B}^*}{q_t^l}. \quad (62)$$

*Proof.* First, We have

$$\begin{aligned} \tilde{V}_n^2(p_t, \nu_t) &= \max_{x \in X} \min_{q \in \Delta(L)} \min_{\tau \in \mathcal{T}} \sum_{l \in L} q^l \left( \nu_t^l + E \left( \sum_{s=1}^n M(k, l, a_s, b_s) | l \right) \right), \\ &= \max_{x \in X} \min_{q \in \Delta(L)} \sum_{l \in L} q^l \left( \nu_t^l + \min_{\tau(l) \in \mathcal{T}(l)} E \left( \sum_{s=1}^n M(k, l, a_s, b_s) | l \right) \right), \\ &= \max_{x \in X} \min_{q \in \Delta(L)} \sum_{l \in L} q^l (\nu_t^l + u_{l,0}(x)). \end{aligned}$$

Define  $\tilde{u}(x) = \min_{q \in \Delta(L)} \sum_{l \in L} q^l (\nu_t^l + u_{l,0}(x))$ . According to the dual theorem, given  $x$ , we have

$$\begin{aligned} \tilde{u}(x) &= \max_{\tilde{u} \in \mathbb{R}} \tilde{u} \\ s.t. \nu_t + u_{:,0}(x) &\geq \tilde{u} \mathbf{1}, \end{aligned}$$

where  $u_{:,0}(x)$  satisfies (29-31) with the horizon to be  $n$ . Therefore, following the same steps in the proof of Theorem 2 to show  $\hat{u} = u$ , we have

$$\begin{aligned} \tilde{u}(x) &= \max_{u \in U, u_{:,0} \in \mathbb{R}^{|L|}, \tilde{u} \in \mathbb{R}} \tilde{u} \\ s.t. \nu_t + u_{:,0} &\geq \tilde{u} \mathbf{1}, \\ \sum_{k \in K} M^{klT} x_{k, h_1^A, h_1^B} + u_{l, h_1^A, h_1^B}^T \mathbf{1} &\geq u_{l,0} \mathbf{1}, \quad \forall l \in L, \\ \sum_{k \in K} M^{klT} x_{k, h_{t+1}^A, h_{t+1}^B} + u_{l, h_{t+1}^A, h_{t+1}^B}^T \mathbf{1} &\geq u_{l, h_t^A, h_t^B}^{a_t, b_t} \mathbf{1}, \quad \forall t = 1, \dots, n-1, l \in L, h_t^A \in H_t^A, h_t^B \in H_t^B. \end{aligned}$$

Hence, equation (53-56) is shown. Player 1's security strategy at stage  $t$  in dual game  $\tilde{\Gamma}_T^2(p, \nu)$  can be seen as player 1's security strategy at stage 1 in dual game  $\tilde{\Gamma}_n^2(p_t, \nu_t)$ . Hence, according to equation (1), we have  $\tilde{\sigma}_t^*(k, p_t, \nu_t) = x_{k, h_1^A, h_1^B}^* / p_t^k$ .

Following the same steps, we show equation (58-61) is also true, and player 2's security strategy at stage  $t$  is as in (62).  $\square$

Now, let's get back to the primal  $T$ -stage repeated Bayesian game  $\Gamma_T(p, q)$ . It was shown in [1], [9], [12] that if  $\nu^*$  is the optimal solution to  $\max_{\nu \in \mathbb{R}^{|\mathcal{L}|}} \{\tilde{V}_T^2(q, \nu) - q^T \nu\}$ , then player 1's security strategy in type 2 dual game  $\tilde{\Gamma}_T^2(p, \nu^*)$  is also the player's security strategy in the primal game  $\Gamma_T(p, q)$ , and that if  $\mu^*$  is the optimal solution to  $\min_{\mu \in \mathbb{R}^{|\mathcal{K}|}} \{\tilde{V}_T^1(\mu, q) - p^T \mu\}$ , then player's security strategy in type 1 dual game  $\tilde{\Gamma}_T^1(\mu^*, q)$  is also the player's security strategy in the primal game  $\Gamma_T(p, q)$ . Since Lemma 3 shows that  $\nu^*$  and  $\mu^*$  are the regrets on player 2 and 1's type, respectively, we have the following corollary.

**Corollary 5.** Consider a  $T$ -stage repeated Bayesian game  $\Gamma_T(p, q)$  and its dual games  $\tilde{\Gamma}_T^1(\mu, q)$  and  $\tilde{\Gamma}_T^2(p, \nu)$ . Player 1's security strategy  $\tilde{\sigma}^* \in \Sigma$ , which depends only on  $t$ ,  $p_t$  and  $\nu_t$  at stage  $t$ , in type 2 dual game  $\tilde{\Gamma}_T^2(p, \nu^*)$  is also player 1's security strategy in the primal game  $\Gamma_T(p, q)$ , where  $\nu^*$  is given in (39).

Similarly, player 2's security strategy  $\tilde{\tau}^* \in \mathcal{T}$ , which depends only on  $t$ ,  $\mu_t$  and  $q_t$  at stage  $t$ , in type 1 dual game  $\tilde{\Gamma}_T^1(\mu^*, q)$  is also player 2's security strategy in the primal game  $\Gamma_T(p, q)$ , where  $\mu^*$  is given in (38).

According to Corollary 5, we can compute player 1's security strategy in the following way. First, compute the initial regret,  $\nu^*$ , on player 2's type. Stage by stage, update  $p_t$  and  $\nu_t$ , and compute the security strategy based on  $p_t$ ,  $\nu_t$  and  $t$  in the dual game. Player 2's security strategy is computed in the same way.

**Algorithm 6.** Player 1's security strategy based on fixed-sized sufficient statistics

- 1) Initialization
  - Compute  $u_{:,0}^*$  based on LP (11-13).
  - Set  $t = 1$ ,  $p_t = p$ , and  $\nu_t = -u_{:,0}^*$ .
- 2) Compute player 1's security strategy  $\tilde{\sigma}_t^*$  at stage  $t$  according to (57) based on LP (53-56).
- 3) Choose an action in  $A$  according to  $\tilde{\sigma}_t^*$ , and announce the action publicly. Meanwhile, read player 2's current action.
- 4) If  $t = T$ , then go to step 6. Otherwise, update  $p_{t+1}$  and  $\nu_{t+1}$  according to (49) and (51), respectively.
- 5) Update  $t = t + 1$  and go to step 2.
- 6) End.

**Algorithm 7.** Player 2's security strategy based on fixed-sized sufficient statistics

- 1) Initialization
  - Compute  $w_{:,0}^*$  based on LP (15-17).
  - Set  $t = 1$ ,  $\mu_t = -w_{:,0}^*$ , and  $q_t = q$ .
- 2) Compute Player 2's security strategy  $\tilde{\tau}_t^*$  at stage  $t$  according to (62) based on LP (58-61).
- 3) Choose an action in  $B$  according to  $\tilde{\tau}_t^*$ , and announce it publicly. Meanwhile, read player 1's current action.
- 4) If  $t = T$ , then go to step 6. Otherwise, update  $q_{t+1}$  and  $\mu_{t+1}$  according to (44) and (46), respectively.
- 5) Update  $t = t + 1$  and go to step 2.
- 6) End.

### III. $\lambda$ -DISCOUNTED REPEATED BAYESIAN GAMES

A two-player zero-sum  $\lambda$ -discounted repeated Bayesian game, which is simply called discounted game or discounted primal game in the rest of this paper, is specified by the same seven-tuple  $(K, L, A, B, M, p_0, q_0)$  and played in the same way as in a two-player zero-sum  $T$ -stage repeated Bayesian game. The payoff of player 1 at stage  $t$  is  $\lambda(1 - \lambda)^{t-1} M(k, l, a_t, b_t)$ , where  $\lambda \in (0, 1)$ , and the game is played for infinite horizon. Correspondingly, the strategy spaces  $\Sigma$  and  $\mathcal{T}$  are defined for infinite horizon. The total payoff of the discounted game with initial probability  $p_0, q_0$  and strategies  $\sigma$  and  $\tau$  is defined as

$$\gamma_\lambda(p_0, q_0, \sigma, \tau) = E_{p_0, q_0, \sigma, \tau} \left( \sum_{t=1}^{\infty} \lambda(1 - \lambda)^{t-1} M(k, l, a_t, b_t) \right).$$

The discounted game  $\Gamma_\lambda(p_0, q_0)$  is defined as a two-player zero-sum repeated Bayesian game equipped with initial distribution  $p_0$  and  $q_0$ , strategy spaces  $\Sigma$  and  $\mathcal{T}$ , and payoff function  $\gamma_\lambda(p_0, q_0, \sigma, \tau)$ . The security strategies  $\sigma^*$  and  $\tau^*$ , and security levels  $\underline{V}_\lambda(p_0, q_0)$  and  $\bar{V}_\lambda(p_0, q_0)$  are defined in the same way as in a  $T$ -stage repeated Bayesian game. Since  $\gamma_\lambda(p_0, q_0, \sigma, \tau)$  is bilinear over  $\sigma$  and  $\tau$ , the discounted game  $\Gamma_\lambda(p_0, q_0)$  has a value  $V_\lambda(p_0, q_0)$  according to Sion's minmax Theorem, i.e.  $V_\lambda(p_0, q_0) = \underline{V}_\lambda(p_0, q_0) = \bar{V}_\lambda(p_0, q_0)$ .

### A. Dual games, security strategies, and sufficient statistics

A discounted game is played for infinite stages, and the history action space is infinite, too. It is not practical to design behavior strategies that directly depends on history actions, and it is necessary to find a sufficient statistics for decision making. A candidate sufficient statistics in the discounted game  $\Gamma_\lambda(p, q)$  is the belief state pair  $(p_t, q_t)$  [9], [1]. The belief state pair is, unfortunately, not fully available to either player after the first stage, since  $(p_t, q_t)$  depends on both players' strategies according to (49) and (44). The objective in this subsection is to find, for every player, the fully available sufficient statistics and the corresponding security strategy that depends on the sufficient statistics. We will use the same technique as that in the  $T$ -stage game to find the fully available sufficient statistics and the corresponding security strategies. Let's start from the dual games of the discounted game.

A discounted game  $\Gamma_\lambda(p, q)$  also has two dual games. The discounted type 1 dual game  $\tilde{\Gamma}_\lambda^1(\mu, q)$  is with respect to the first parameter  $p$ , where  $\mu \in \mathbb{R}^{|\mathcal{K}|}$  is the initial regret with respect to player 1's type. The discounted type 1 dual game  $\tilde{\Gamma}_\lambda^1(\mu, q)$  is played the same as in the  $T$ -stage type 1 dual game  $\tilde{\Gamma}_T^1(\mu, q)$ , except that the discounted game is played for infinite horizon. Let  $p$  be player 1's strategy to choose his own type. Player 1's payoff is

$$\tilde{\gamma}_\lambda^1(\mu, q, p, \sigma, \tau) = E_{p, q, \sigma, \tau} \left( \mu^k + \sum_{t=1}^{\infty} \lambda(1 - \lambda)^{t-1} M(k, l, a_t, b_t) \right).$$

The discounted type 2 dual game  $\tilde{\Gamma}_\lambda^2(p, \nu)$  is defined with respect to the second parameter  $q$ , where  $\nu \in \mathbb{R}^{|\mathcal{L}|}$  is the initial regret with respect to player 2's type. The discounted type 2 dual game  $\tilde{\Gamma}_\lambda^2(p, \nu)$  is played the same as in the  $T$ -stage type 2 dual game  $\tilde{\Gamma}_T^2(p, \nu)$ , except that the discounted game is played for infinite horizon. Let  $q$  be player 2's strategy to choose his type. Player 1's payoff is

$$\tilde{\gamma}_\lambda^2(p, \nu, q, \sigma, \tau) = E_{p, q, \sigma, \tau} \left( \nu^l + \sum_{t=1}^{\infty} \lambda(1 - \lambda)^{t-1} M(k, l, a_t, b_t) \right).$$

Both dual games,  $\tilde{\Gamma}_\lambda^1(\mu, q)$  and  $\tilde{\Gamma}_\lambda^2(p, \nu)$ , have values indicated by  $\tilde{V}_\lambda^1(\mu, q)$  and  $\tilde{V}_\lambda^2(p, \nu)$ , respectively. The game values of the discounted dual games are related to the game value of the discounted primal game in the following way [9].

$$\tilde{V}_\lambda^1(\mu, q) = \max_{p \in \Delta(\mathcal{K})} \{V_\lambda(p, q) + p^T \mu\} \quad (63)$$

$$V_\lambda(p, q) = \min_{\mu \in \mathbb{R}^{|\mathcal{K}|}} \{\tilde{V}_\lambda^1(\mu, p) - p^T \mu\} \quad (64)$$

$$\tilde{V}_\lambda^2(p, \nu) = \min_{q \in \Delta(\mathcal{L})} \{V_\lambda(p, q) + q^T \nu\} \quad (65)$$

$$V_\lambda(p, q) = \max_{\nu \in \mathbb{R}^{|\mathcal{L}|}} \{\tilde{V}_\lambda^2(p, \nu) - q^T \nu\} \quad (66)$$

Let  $\mu^*$  and  $\nu^*$  be the optimal solution to the optimal problem on the right hand side of (64) and (66), respectively. Player 2's security strategy in discounted type 1 dual game  $\tilde{\Gamma}_\lambda^1(\mu^*, q)$  is also his security strategy in the discounted primal game  $\Gamma_\lambda(p, q)$  [9], [12], [1]. Player 1's security strategy in discounted type 2 dual game  $\tilde{\Gamma}_\lambda^2(p, \nu^*)$  is his security strategy in the discounted primal game  $\Gamma_\lambda(p, q)$  [9]. The following lemma tells us what  $\mu^*$  and  $\nu^*$  are. The proof is the same as the proof of Lemma 3

**Lemma 8.** Consider a discounted game  $\Gamma_\lambda(p, q)$ . Let  $\sigma_{p,q}^*$  and  $\tau_{p,q}^*$  be player 1 and 2's security strategies, and  $x_{p,q}^*$  and  $y_{p,q}^*$  be the corresponding optimal realization plans of player 1 and 2. Define

$$u_{l,0;\lambda}(x) = \min_{\tau(l) \in \mathcal{T}(l)} E \left( \sum_{t=1}^{\infty} \lambda(1 - \lambda)^{t-1} M(k, l, a_t, b_t) | l \right),$$

$$w_{k,0;\lambda}(y) = \max_{\sigma(k) \in \Sigma(k)} E \left( \sum_{t=1}^{\infty} \lambda(1 - \lambda)^{t-1} M(k, l, a_t, b_t) | k \right),$$

where  $x$  and  $y$  are player 1 and 2's realization plans. The optimal solution  $\mu^*$  to the optimal problem  $\min_{\mu \in \mathbb{R}^{|\mathcal{K}|}} \{\tilde{V}_\lambda^1(\mu, q) - p^T \mu\}$  is

$$\mu^{*k} = -w_{k,0;\lambda}(y_{p,q}^*), \forall k \in \mathcal{K}. \quad (67)$$

The optimal solution  $\nu^*$  to the optimal problem  $\max_{\nu \in \mathbb{R}^{|\mathcal{L}|}} \{\tilde{V}_\lambda^2(p, \nu) - q^T \nu\}$  is

$$\nu^{*l} = -u_{l,0;\lambda}(x_{p,q}^*), \forall l \in \mathcal{L}. \quad (68)$$

Now that we've found the special initial regrets in the dual games, our next step is to study players' security strategies in the dual games. With a little abuse of notation  $\mu_t$  and  $\nu_t$ , in discounted games,  $\mu_t$  and  $\nu_t$  are called the *anti-discounted regret*

on player 1 and 2's types, respectively. In discounted type 1 dual game  $\tilde{\Gamma}_\lambda^1(\mu, q)$ , the anti-discounted regret  $\mu_t$  on player 1's type is defined as

$$\mu_t^k = \frac{\mu^k + \sum_{s=1}^{t-1} \lambda(1-\lambda)^{s-1} E(M(k, l, a_s, b_s) | k, h_{s+1}^A, h_{s+1}^B)}{(1-\lambda)^{t-1}}, \forall k \in K, t = 1, 2, \dots, \quad (69)$$

and is updated as follows

$$\mu_{t+1}^k = \mu^+( \mu_t, a_t, b_t, z_t, q_t ) = \frac{\mu_t^k + \lambda \sum_{l \in L} q_{t+1}^l M_{a_t, b_t}^{kl}}{1-\lambda}, \forall k \in K, \text{ with } \mu_1 = \mu, \quad (70)$$

where  $q_t$  is the belief on player 2's type defined as in (44), and updated as in (44).

In discounted type 2 dual game  $\tilde{\Gamma}_\lambda^2(p, \nu)$ , the anti-discounted regret  $\nu_t$  on player 2's type is defined as

$$\nu_t^l = \frac{\nu^l + \sum_{s=1}^{t-1} \lambda(1-\lambda)^{s-1} E(M(k, l, a_s, b_s) | l, h_{s+1}^A, h_{s+1}^B)}{(1-\lambda)^{t-1}}, \forall l \in L, t = 1, 2, \dots, \quad (71)$$

and is updated as follows

$$\nu_{t+1}^l = \nu^+( \nu_t, a_t, b_t, r_t, p_t ) = \frac{\nu_t^l + \lambda \sum_{k \in K} p_{t+1}^k M_{a_t, b_t}^{kl}}{1-\lambda}, \forall l \in L, \text{ with } \nu_1 = \nu, \quad (72)$$

where  $p_t$  is the belief on player 1's type defined as in (49), and updated as in (49).

Player 2's security strategy in discounted type 1 dual game can be found by solving equation (73), and depends only on  $\mu_t$  and  $q_t$  at stage  $t$ . Player 1's security strategy in discounted type 2 dual game can be computed by solving equation (74), and depends only on  $q_t$  and  $\nu_t$  at stage  $t$ . The applaudable property of  $(\mu_t, q_t)$  and  $(p_t, \nu_t)$  is that they are fully available to player 2 and 1, respectively.

$$\tilde{V}_\lambda^1(\mu, q) = \min_{z \in \Delta(B)^{|L|}} \max_{a \in A} (1-\lambda) \sum_{b \in B} \bar{z}_{q,z}(b) \tilde{V}_\lambda^1(\mu^+( \mu, a, b, z, q ), q^+(b, z, q)). \quad (73)$$

$$\tilde{V}_\lambda^2(p, \nu) = \max_{r \in \Delta(A)^{|K|}} \min_{b \in B} (1-\lambda) \sum_{a \in A} \bar{r}_{p,r}(a) \tilde{V}_\lambda^2(p^+(a, r, p), \nu^+( \nu, a, b, r, p)). \quad (74)$$

Finally, we are ready to investigate players' sufficient statistics and the corresponding security strategies in the discounted primal game  $\Gamma_\lambda(p, q)$ .

**Corollary 9.** [9] Consider a discounted game  $\Gamma_\lambda(p, q)$  and its dual games  $\tilde{\Gamma}_\lambda^1(\mu, q)$  and  $\tilde{\Gamma}_\lambda^2(p, \nu)$ . Let  $\nu^*$  take the form of (68). Player 1's security strategy  $\tilde{\sigma}^*$ , which depends only on  $(p_t, \nu_t)$  at stage  $t$ , in discounted type 2 dual game  $\tilde{\Gamma}_\lambda^2(p, \nu^*)$  is also player 1's security strategy in the discounted primal game  $\Gamma_\lambda(p, q)$ .

Similarly, let  $\mu^*$  take the form of (67). Player 2's security strategy  $\tilde{\tau}^*$ , which depends only on  $(\mu_t, q_t)$  at stage  $t$ , in discounted type 1 dual game  $\tilde{\Gamma}_\lambda^1(\mu^*, q)$  is also player 2's security strategy in the discounted primal game  $\Gamma_\lambda(p, q)$ .

### B. Approximate the initial regret states $\mu^*$ and $\nu^*$

To compute players' security strategies in the discounted primal game, the first thing is to compute the initial regrets,  $\mu^* = -w_{:,0;\lambda}(y^*)$  and  $\nu^* = -u_{:,0;\lambda}(x^*)$ , which is non-convex [13]. Therefore, we consider using the game value of a  $\lambda$ -discounted  $T$ -stage game to approximate the game value of the discounted game with infinite horizon, and further find approximated  $\mu^*$  and  $\nu^*$ .

A  $\lambda$ -discounted  $T$ -stage repeated Bayesian game  $\Gamma_{\lambda,T}(p, q)$  is a truncated version of the  $\lambda$ -discounted repeated Bayesian game  $\Gamma_\lambda(p, q)$  with the time horizon to be  $T$ . We denote the payoff and the game value of the truncated discounted game as  $\gamma_{\lambda,T}(p, q, \sigma, \tau)$  and  $V_{\lambda,T}(p, q)$ , respectively. In game  $\Gamma_{\lambda,T}(p, q)$ , we define anti-discounted weighted future security payoffs  $u_{l, h_t^A, h_t^B; \lambda, T}^{a_t, b_t}$  and  $w_{k, h_t^A, h_t^B; \lambda, T}^{a_t, b_t}$  for  $t = 0, \dots, T-1$  as follows

$$u_{l, h_t^A, h_t^B; \lambda, T}^{a_t, b_t}(x) = (1-\lambda)^{-t} \min_{\tau_{t+1:T}(l) \in \mathcal{T}_{t+1:T}(l)} \sum_{k \in K} x_{k, h_t^A, h_t^B}^{a_t} E \left( \sum_{s=t+1}^T \lambda(1-\lambda)^{s-1} M(k, l, a_s, b_s) | k, l, h_{t+1}^A, h_{t+1}^B \right), \quad (75)$$

$$w_{k, h_t^A, h_t^B; \lambda, T}^{a_t, b_t}(y) = (1-\lambda)^{-t} \max_{\sigma_{t+1:T}(k) \in \Sigma_{t+1:T}(k)} \sum_{l \in L} y_{l, h_t^A, h_t^B}^{b_t} E \left( \sum_{s=t+1}^T \lambda(1-\lambda)^{s-1} M(k, l, a_s, b_s) | k, l, h_{t+1}^A, h_{t+1}^B \right), \quad (76)$$

where  $h_{t+1}^A = (h_t^A, a_t)$  and  $h_{t+1}^B = (h_t^B, b_t)$ , and the pairs here indicate concatenation.

Notice that  $u_{l, h_0^A, h_0^B; \lambda, T}^{a_0, b_0}(x)$  and  $w_{k, h_0^A, h_0^B; \lambda, T}^{a_0, b_0}(y)$ , also denoted as  $u_{l,0;\lambda,T}(x)$  and  $w_{k,0;\lambda,T}(y)$ , are truncated versions of  $u_{l,0;\lambda}(x)$  and  $w_{k,0;\lambda}(y)$ , respectively. We can use  $u_{l,0;\lambda,T}(x^*)$  and  $w_{k,0;\lambda,T}(y^*)$  to approximate  $u_{l,0;\lambda}(x^*)$  and  $w_{k,0;\lambda}(y^*)$ , and hence  $\nu^*$  and  $\mu^*$ . Here,  $x^*$  and  $y^*$  are player 1 and 2's optimal realization plan in game  $\Gamma_\lambda(p, q)$ , and  $x^*$  and  $y^*$  are player 1

and 2's security strategies in game  $\Gamma_{\lambda,T}(p, q)$ . The following theorem provides linear programs to compute  $u_{l,0;\lambda,T}(x^*)$  and  $w_{k,0;\lambda,T}(y^*)$ .

**Theorem 10.** Consider a  $\lambda$ -discounted  $T$ -stage repeated Bayesian game  $\Gamma_{\lambda,T}(p, q)$ . Its game value  $V_{\lambda,T}(p, q)$  satisfies

$$V_{\lambda,T}(p, q) = \max_{x \in X, u_{\lambda,T} \in U, u_{:,0;\lambda,T} \in \mathbb{R}^{|L|}} \sum q^l u_{l,0;\lambda,T} \quad (77)$$

$$s.t. \lambda \sum_{k \in K} M^{kl} x_{k,h_1^A,h_1^B} + (1-\lambda) u_{l,h_1^A,h_1^B;\lambda,T} \mathbf{1} \geq u_{l,0;\lambda,T} \mathbf{1}, \quad \forall l \in L, \quad (78)$$

$$\lambda \sum_{k \in K} M^{kl} x_{k,h_{t+1}^A,h_{t+1}^B} + (1-\lambda) u_{l,h_{t+1}^A,h_{t+1}^B;\lambda,T} \mathbf{1} \geq u_{l,h_t^A,h_t^B;\lambda,T}^{a_t,b_t} \mathbf{1}, \quad (79)$$

$$\forall t = 1, \dots, T-1, l \in L, h_{t+1} \in H_{t+1}^A, h_{t+1}^B \in H_{t+1}^B,$$

where  $h_{t+1}^A = (h_t^A, a_t)$ ,  $h_{t+1}^B = (h_t^B, b_t)$ ,  $X$  is a set including all real vectors satisfying (3-4), and  $U$  is a real space of an appropriate dimension. The optimal solution  $x_{\lambda,T}^*$  is player 1's optimal realization plan in game  $\Gamma_{\lambda,T}(p, q)$ , and its anti-discounted weighted future security payoff at stage 0 is the optimal solution  $u_{:,0;\lambda,T}^*$ , i.e.  $u_{:,0;\lambda,T}(x^*) = u_{:,0;\lambda,T}^*$ .

Dually,  $V_{\lambda,T}(p, q)$  also satisfies

$$V_{\lambda,T}(p, q) = \min_{y \in Y, w_{\lambda,T} \in W, w_{:,0;\lambda,T} \in \mathbb{R}^{|K|}} \sum_{k \in K} p^k w_{k,0;\lambda,T} \quad (80)$$

$$s.t. \lambda \sum_{l \in L} M^{kl} y_{l,h_1^A,h_1^B} + (1-\lambda) w_{k,h_1^A,h_1^B;\lambda,T} \mathbf{1} \leq w_{k,0;\lambda,T} \mathbf{1}, \quad \forall k \in K, \quad (81)$$

$$\lambda \sum_{l \in L} M^{kl} y_{l,h_{t+1}^A,h_{t+1}^B} + (1-\lambda) w_{k,h_{t+1}^A,h_{t+1}^B;\lambda,T} \mathbf{1} \leq w_{k,h_t^A,h_t^B;\lambda,T}^{a_t,b_t} \mathbf{1}, \quad (82)$$

$$\forall t = 1, \dots, T-1, k \in K, h_{t+1}^A \in H_{t+1}^A, h_{t+1}^B \in H_{t+1}^B,$$

where  $h_{t+1}^A = (h_t^A, a_t)$ ,  $h_{t+1}^B = (h_t^B, b_t)$ ,  $Y$  is a set including all real vectors satisfying (5-6), and  $W$  is a real space of an appropriate dimension. The optimal solution  $y^*$  is player 2's optimal realization plan in game  $\Gamma_{\lambda,T}(p, q)$ , and its anti-discounted weighted future security payoff at stage 0 is the optimal solution  $w_{:,0;\lambda,T}^*$ , i.e.  $w_{:,0;\lambda,T}(y^*) = w_{:,0;\lambda,T}^*$ .

*Proof.* Following the same steps as in the proof of Lemma 1, we have for all  $t = 0, \dots, T-1$ ,

$$u_{l,h_t^A,h_t^B;\lambda,T}^{a_t,b_t}(x) = \min_{\tau_{t+1}(l, h_{t+1}^A, h_{t+1}^B) \in \Delta(B)} \left( \lambda \sum_{k \in K} x_{k,h_{t+1}^A,h_{t+1}^B}^T M^{kl} + (1-\lambda) \mathbf{1}^T u_{l,h_{t+1}^A,h_{t+1}^B;\lambda,T}(x) \right) \tau_{t+1}(l, h_{t+1}^A, h_{t+1}^B), \quad (83)$$

$$w_{k,h_t^A,h_t^B;\lambda,T}^{a_t,b_t}(y) = \max_{\sigma_{t+1}(k, h_{t+1}^A, h_{t+1}^B) \in \Delta(A)} \sigma_{t+1}(k, h_{t+1}^A, h_{t+1}^B)^T \left( \lambda \sum_{l \in L} M^{kl} y_{l,h_{t+1}^A,h_{t+1}^B} + (1-\lambda) w_{k,h_{t+1}^A,h_{t+1}^B;\lambda,T}(y) \mathbf{1} \right). \quad (84)$$

Following the same steps as in the proof of Theorem 2, we show Theorem 10 is true.  $\square$

With  $u^*(:, 0; \lambda, T)$  and  $w^*(:, 0; \lambda, T)$  computed based on (77-79) and (80-82), according to Lemma 8, we approximate  $\mu^*$  and  $\nu^*$  as

$$\mu^\dagger = -w^*(:, 0; \lambda, T), \text{ and} \quad (85)$$

$$\nu^\dagger = -u^*(:, 0; \lambda, T), \quad (86)$$

respectively.

### C. Approximate the security strategies $\tilde{\sigma}^*$ and $\tilde{\tau}^*$ in dual games

Now that we have constructed an LP to compute the approximated initial regrets  $\mu^\dagger$  and  $\nu^\dagger$  in the dual games, the next step is to compute the security strategy in a discounted dual game, which will serve as the security strategy of the corresponding player in the discounted primal game.

Computing the security strategies and the game values in dual games  $\tilde{\Gamma}_\lambda^1(\mu^*, q)$  and  $\tilde{\Gamma}_\lambda^2(p, \nu^*)$  is non-convex [13]. Therefore, we use the game values of truncated discounted dual games to approximate the game value of the discounted dual games, and then compute approximated security strategies based on the approximated game value.

A  $\lambda$ -discounted  $T$ -stage type 1 dual game  $\tilde{\Gamma}_{\lambda,T}^1(\mu, q)$  is a truncated discounted type 1 dual game  $\tilde{V}_\lambda^1(\mu, q)$  with time horizon to be  $T$  stages. Following the same step as in the proof of Proposition 4.22 in [9], the game value  $\tilde{V}_{\lambda,T+1}^1(\mu, q)$  of the  $\lambda$ -discounted  $T+1$ -stage type 1 dual game  $\tilde{\Gamma}_{\lambda,T+1}^1(\mu, q)$  satisfies

$$\tilde{V}_{\lambda,T+1}^1(\mu, q) = \min_{z \in \Delta(B)^{|L|}} \max_{a \in A} (1-\lambda) \sum_{b \in B} \tilde{z}_{q,z}(b) \tilde{V}_{\lambda,T}^1(\mu^+( \mu, a, b, z, q), q^+(b, z, q)), \quad (87)$$

with  $\tilde{V}_{\lambda,0}^1(\mu, q) = \max\{\mu\}$ . Moreover, since  $\tilde{\Gamma}_{\lambda,T}^1(\mu, q)$  is a dual game of  $\Gamma_{\lambda,T}(p, q)$  with respect to the first parameter  $p$ , their game values satisfy

$$\tilde{V}_{\lambda,T}^1(\mu, q) = \max_{p \in \Delta(K)} \{V_{\lambda,T}(p, q) + p^T \mu\}, \quad (88)$$

$$V_{\lambda,T}(p, q) = \min_{\mu \in \mathbb{R}^{|K|}} \{\tilde{V}_{\lambda,T}^1(\mu, q) - p^T \mu\}. \quad (89)$$

Similarly, a type 2  $\lambda$ -discounted  $T$ -stage dual game  $\tilde{\Gamma}_{\lambda,T}^2(p, \nu)$  is the truncated version of discounted type 2 dual game with time horizon  $T$ . The game value  $\tilde{V}_{\lambda,T+1}^2(p, \nu)$  of the truncated discounted type 2 dual game  $\tilde{\Gamma}_{\lambda,T+1}^2(p, \nu)$  satisfies

$$\tilde{V}_{\lambda,T+1}^2(p, \nu) = \max_{r \in \Delta(A)^{|K|}} \min_{b \in B} (1 - \lambda) \sum_{a \in A} \bar{r}_{p,r}(a) \tilde{V}_{\lambda,T}^2(p^+(a, r, p), \nu^+(\nu, a, b, r, p)), \quad (90)$$

with  $\tilde{V}_{\lambda,0}^2(p, \nu) = \min\{\nu\}$ . The truncated discounted type 2 dual game  $\tilde{\Gamma}_{\lambda,T}^2(p, \nu)$  is the dual game of the truncated discounted game  $\Gamma_{\lambda,T}(p, q)$  with respect to the second parameter  $q$ , and hence their game values satisfy

$$\tilde{V}_{\lambda,T}^2(p, \nu) = \min_{q \in \Delta(L)} \{V_{\lambda,T}(p, q) + q^T \nu\}, \quad (91)$$

$$V_{\lambda,T}(p, q) = \max_{\nu \in \mathbb{R}^{|L|}} \{\tilde{V}_{\lambda,T}^2(p, \nu) - q^T \nu\}. \quad (92)$$

Based on the relations between the game values of the discounted game, the truncated discounted game  $\Gamma_{\lambda,T}(p, q)$ , and their dual games, we have the following lemma.

**Lemma 11.** *Consider a discounted game  $\Gamma_{\lambda}(p, q)$  and its dual games  $\tilde{\Gamma}_{\lambda}^1(\mu, q)$  and  $\tilde{\Gamma}_{\lambda}^2(p, \nu)$ , and a  $T$ -stage discounted game  $\Gamma_{\lambda,T}(p, q)$  and its dual games  $\tilde{\Gamma}_{\lambda,T}^1(\mu, q)$  and  $\tilde{\Gamma}_{\lambda,T}^2(p, \nu)$ . Their game values satisfy*

$$\|V_{\lambda} - V_{\lambda,T}\|_{\sup} = \|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup} = \|\tilde{V}_{\lambda}^2 - \tilde{V}_{\lambda,T}^2\|_{\sup}. \quad (93)$$

*Proof.* First, we show that  $\|V_{\lambda} - V_{\lambda,T}\|_{\sup} \leq \|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup}$ . According to equation (64) and (89), we have

$$|V_{\lambda}(p, q) - V_{\lambda,T}(p, q)| = |\min_{\mu \in \mathbb{R}} \{\tilde{V}_{\lambda}^1(\mu, q) - p^T \mu\} - \min_{\mu \in \mathbb{R}} \{\tilde{V}_{\lambda,T}^1(\mu, q) - p^T \mu\}|.$$

Let  $\mu^*$  and  $\mu^*$  be the optimal solutions to the optimal problem  $\min_{\mu \in \mathbb{R}} \{\tilde{V}_{\lambda}^1(\mu, q) - p^T \mu\}$  and  $\min_{\mu \in \mathbb{R}} \{\tilde{V}_{\lambda,T}^1(\mu, q) - p^T \mu\}$ , respectively. If  $\min_{\mu \in \mathbb{R}} \{\tilde{V}_{\lambda}^1(\mu, q) - p^T \mu\} \geq \min_{\mu \in \mathbb{R}} \{\tilde{V}_{\lambda,T}^1(\mu, q) - p^T \mu\}$ , then we have  $|V_{\lambda}(p, q) - V_{\lambda,T}(p, q)| \leq |\tilde{V}_{\lambda}^1(\mu^*, q) - \tilde{V}_{\lambda,T}^1(\mu^*, q)|$ . Otherwise, it is true that  $|V_{\lambda}(p, q) - V_{\lambda,T}(p, q)| \leq |\tilde{V}_{\lambda}^1(\mu^*, q) - \tilde{V}_{\lambda,T}^1(\mu^*, q)|$ . Therefore, we have, for any  $p \in \Delta(K)$  and  $q \in \Delta(L)$ ,  $|V_{\lambda}(p, q) - V_{\lambda,T}(p, q)| \leq \|\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T}^1(\mu, q)\|_{\sup}$ , which implies that  $\|V_{\lambda} - V_{\lambda,T}\|_{\sup} \leq \|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup}$ .

Next, we show that  $\|V_{\lambda} - V_{\lambda,T}\|_{\sup} \geq \|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup}$ . According to equation (63) and (88), we have

$$|\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T}^1(\mu, q)| = |\max_{p \in \Delta(K)} \{V_{\lambda}(p, q) + p^T \mu\} - \max_{p \in \Delta(K)} \{V_{\lambda,T}(p, q) + p^T \mu\}|.$$

Let  $p^*$  and  $p^*$  be the optimal solutions to the optimal problems  $\max_{p \in \Delta(K)} \{V_{\lambda}(p, q) + p^T \mu\}$  and  $\max_{p \in \Delta(K)} \{V_{\lambda,T}(p, q) + p^T \mu\}$ , respectively. If  $\max_{p \in \Delta(K)} \{V_{\lambda}(p, q) + p^T \mu\} \geq \max_{p \in \Delta(K)} \{V_{\lambda,T}(p, q) + p^T \mu\}$ , then we have  $|\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T}^1(\mu, q)| \leq |V_{\lambda}(p^*, q) - V_{\lambda,T}(p^*, q)|$ . Otherwise, it is true that  $|\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T}^1(\mu, q)| \leq |V_{\lambda}(p^*, q) - V_{\lambda,T}(p^*, q)|$ . Therefore, we conclude that for any  $\mu \in \mathbb{R}^{|K|}$  and  $q \in \Delta(L)$ ,  $|\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T}^1(\mu, q)| \leq \|V_{\lambda}(p, q) - V_{\lambda,T}(p, q)\|_{\sup}$ , which implies that  $\|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup} \leq \|V_{\lambda} - V_{\lambda,T}\|_{\sup}$ .

Therefore, we prove the first equality of (93). Following the same steps, we can show  $\|V_{\lambda} - V_{\lambda,T}\|_{\sup} = \|\tilde{V}_{\lambda}^2 - \tilde{V}_{\lambda,T}^2\|_{\sup}$ .  $\square$

When we use  $\tilde{V}_{\lambda,T}^1(\mu, q)$  and  $\tilde{V}_{\lambda,T}^2(p, \nu)$  to approximate  $\tilde{V}_{\lambda}^1(\mu, q)$  and  $\tilde{V}_{\lambda}^2(p, \nu)$ , respectively, we are interested in how fast the approximations converge to the real game values. To this purpose, we define two operators  $\tilde{F}^1$  and  $\tilde{F}^2$  as

$$\tilde{F}_z^{1, \tilde{V}^1}(\mu, q) = \max_{a \in A} (1 - \lambda) \sum_{b \in B} \bar{z}_{q,z}(b) \tilde{V}^1(\mu^+(\mu, a, b, z, q), q^+(b, z, q)), \quad (94)$$

$$\tilde{F}_r^{2, \tilde{V}^2}(p, \nu) = \min_{b \in B} (1 - \lambda) \sum_{a \in A} \bar{r}_{p,r}(a) \tilde{V}^2(p^+(a, r, p), \nu^+(\nu, a, b, r, p)). \quad (95)$$

The two operators  $\tilde{F}^1$  and  $\tilde{F}^2$  are contraction mappings.

**Lemma 12.** *Given any  $z \in \Delta(K)$ ,  $r \in \Delta(L)$  and  $\lambda \in (0, 1)$ , the operators  $\tilde{F}^1$  and  $\tilde{F}^2$  defined in (94) and (95) are contraction mappings with contraction constant  $1 - \lambda$ , i.e.*

$$\|\tilde{F}_z^{1, \tilde{V}^1} - \tilde{F}_z^{1, \tilde{V}^1}\|_{\sup} \leq (1 - \lambda) \|\tilde{V}_1^1 - \tilde{V}_2^1\|_{\sup}, \quad (96)$$

$$\|\tilde{F}_r^{2, \tilde{V}^2} - \tilde{F}_r^{2, \tilde{V}^2}\|_{\sup} \leq (1 - \lambda) \|\tilde{V}_1^2 - \tilde{V}_2^2\|_{\sup}, \quad (97)$$

where  $\tilde{V}_{1,2}^1 : \mathbb{R}^{|K|} \times \Delta(L) \rightarrow \mathbb{R}$  and  $\tilde{V}_{1,2}^2 : \Delta(K) \times \mathbb{R}^{|L|} \rightarrow \mathbb{R}$ .

*Proof.* Let  $a^*$  and  $a^*$  be the optimal solutions to the optimal problems  $\max_{a \in A} (1-\lambda) \sum_{b \in B} \bar{z}_{q,z} \tilde{V}_1^1(\mu^+(\mu, a, b, z, q), q^+(b, z, q))$  and  $\max_{a \in A} (1-\lambda) \sum_{b \in B} \bar{z}_{q,z} \tilde{V}_2^1(\mu^+(\mu, a, b, z, q), q^+(b, z, q))$ .

If  $\tilde{F}_z^{1, \tilde{V}_1^1}(\mu, q) \geq \tilde{F}_z^{1, \tilde{V}_2^1}(\mu, q)$ , we have

$$\begin{aligned} & |\tilde{F}_z^{1, \tilde{V}_1^1}(\mu, q) - \tilde{F}_z^{1, \tilde{V}_2^1}(\mu, q)| \\ & \leq (1-\lambda) \sum_{b \in B} \bar{z}_{q,z}(b) |\tilde{V}_1^1(\mu^+(\mu, a^*, b, z, q), q^+(b, z, q)) - \tilde{V}_2^1(\mu^+(\mu, a^*, b, z, q), q^+(b, z, q))| \\ & \leq (1-\lambda) \sum_{b \in B} \bar{z}_{q,z}(b) \|\tilde{V}_1^1 - \tilde{V}_2^1\|_{\sup} \\ & = (1-\lambda) \|\tilde{V}_1^1 - \tilde{V}_2^1\|_{\sup}. \end{aligned}$$

Otherwise, we have

$$\begin{aligned} & |\tilde{F}_z^{1, \tilde{V}_1^1}(\mu, q) - \tilde{F}_z^{1, \tilde{V}_2^1}(\mu, q)| \\ & \leq (1-\lambda) \sum_{b \in B} \bar{z}_{q,z}(b) |\tilde{V}_1^1(\mu^+(\mu, a^*, b, z, q), q^+(b, z, q)) - \tilde{V}_2^1(\mu^+(\mu, a^*, b, z, q), q^+(b, z, q))| \\ & \leq (1-\lambda) \sum_{b \in B} \bar{z}_{q,z}(b) \|\tilde{V}_1^1 - \tilde{V}_2^1\|_{\sup} \\ & = (1-\lambda) \|\tilde{V}_1^1 - \tilde{V}_2^1\|_{\sup}. \end{aligned}$$

Hence, for any  $\mu \in \mathbb{R}^{|K|}$  and  $q \in \Delta(L)$ ,  $|\tilde{F}_z^{1, \tilde{V}_1^1}(\mu, q) - \tilde{F}_z^{1, \tilde{V}_2^1}(\mu, q)| \leq (1-\lambda) \|\tilde{V}_1^1 - \tilde{V}_2^1\|_{\sup}$ , which implies equation (96). Equation (97) can be shown following the same steps.  $\square$

Lemma 12 further implies that our game value approximations  $\tilde{V}_{\lambda,T}^1(\mu, q)$  and  $\tilde{V}_{\lambda,T}^2(p, \nu)$  converge to the real game values  $\tilde{V}_{\lambda}^1(\mu, q)$  and  $\tilde{V}_{\lambda}^2(p, \nu)$  exponentially fast with respect to stage  $T$ .

**Theorem 13.** Consider the  $\lambda$ -discounted repeated Bayesian dual games  $\tilde{\Gamma}_{\lambda}^1(\mu, q)$  and  $\tilde{\Gamma}_{\lambda}^2(p, \nu)$ , and their game values  $\tilde{V}_{\lambda}^1(\mu, q)$  and  $\tilde{V}_{\lambda}^2(p, \nu)$ . The game values  $\tilde{V}_{\lambda,T}^1(\mu, q)$  and  $\tilde{V}_{\lambda,T}^2(p, \nu)$  of  $\lambda$ -discounted  $T$ -stage dual games  $\tilde{\Gamma}_{\lambda,T}^1(\mu, q)$  and  $\tilde{\Gamma}_{\lambda,T}^2(p, \nu)$  converge to  $\tilde{V}_{\lambda}^1(\mu, q)$  and  $\tilde{V}_{\lambda}^2(p, \nu)$  exponentially fast with respect to the time horizon  $T$  with convergence rate  $1-\lambda$ , i.e.

$$\|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T+1}^1\|_{\sup} \leq (1-\lambda) \|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup}, \quad (98)$$

$$\|\tilde{V}_{\lambda}^2 - \tilde{V}_{\lambda,T+1}^2\|_{\sup} \leq (1-\lambda) \|\tilde{V}_{\lambda}^2 - \tilde{V}_{\lambda,T}^2\|_{\sup}. \quad (99)$$

*Proof.* Equation (73) and (87) and the definition of  $\tilde{F}^1$  in (94) imply that

$$|\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T+1}^1(\mu, q)| = \left| \min_{z \in \Delta(L)} \tilde{F}_z^{1, \tilde{V}_{\lambda}^1}(\mu, q) - \min_{z \in \Delta(L)} \tilde{F}_z^{1, \tilde{V}_{\lambda,T}^1}(\mu, q) \right|.$$

Let  $z^*$  and  $z^*$  be the optimal solutions to the optimal problems  $\min_{z \in \Delta(L)} \tilde{F}_z^{1, \tilde{V}_{\lambda}^1}(\mu, q)$  and  $\min_{z \in \Delta(L)} \tilde{F}_z^{1, \tilde{V}_{\lambda,T}^1}(\mu, q)$ , respectively. If  $\min_{z \in \Delta(L)} \tilde{F}_z^{1, \tilde{V}_{\lambda}^1}(\mu, q) \geq \min_{z \in \Delta(L)} \tilde{F}_z^{1, \tilde{V}_{\lambda,T}^1}(\mu, q)$ , according to equation (96), we have

$$\begin{aligned} & |\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T+1}^1(\mu, q)| \leq |\tilde{F}_{z^*}^{1, \tilde{V}_{\lambda}^1}(\mu, q) - \tilde{F}_{z^*}^{1, \tilde{V}_{\lambda,T}^1}(\mu, q)| \\ & \leq \|\tilde{F}_{z^*}^{1, \tilde{V}_{\lambda}^1} - \tilde{F}_{z^*}^{1, \tilde{V}_{\lambda,T}^1}\|_{\sup} \leq (1-\lambda) \|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup}. \end{aligned}$$

Otherwise, we have

$$\begin{aligned} & |\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T+1}^1(\mu, q)| \leq |\tilde{F}_{z^*}^{1, \tilde{V}_{\lambda}^1}(\mu, q) - \tilde{F}_{z^*}^{1, \tilde{V}_{\lambda,T}^1}(\mu, q)| \\ & \leq \|\tilde{F}_{z^*}^{1, \tilde{V}_{\lambda}^1} - \tilde{F}_{z^*}^{1, \tilde{V}_{\lambda,T}^1}\|_{\sup} \leq (1-\lambda) \|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup}. \end{aligned}$$

Hence, for any  $\mu \in \mathbb{R}^{|K|}$  and  $q \in \Delta(L)$ ,  $|\tilde{V}_{\lambda}^1(\mu, q) - \tilde{V}_{\lambda,T+1}^1(\mu, q)| \leq (1-\lambda) \|\tilde{V}_{\lambda}^1 - \tilde{V}_{\lambda,T}^1\|_{\sup}$ , which implies equation (98). Equation (99) can be shown following the same steps.  $\square$

With the approximated values  $\tilde{V}_{\lambda,T}^1(\mu, q)$  and  $\tilde{V}_{\lambda,T}^2(p, \nu)$ , we can use them in equation (73) and (74), and derive player 1 and 2's approximated security strategies  $\tilde{\sigma}^\dagger$  and  $\tilde{\tau}^\dagger$  as

$$\tilde{\sigma}^\dagger(:, p, \nu) = \arg \max_{r \in \Delta(A)^{|K|}} \min_{b \in B} (1-\lambda) \sum_{a \in A} \bar{r}_{p,r}(a) \tilde{V}_{\lambda,T}^2(p^+(a, r, p), \nu^+(\nu, a, b, r, p)), \quad (100)$$

$$\tilde{\tau}^\dagger(:, \mu, q) = \arg \min_{z \in \Delta(B)^{|L|}} \max_{a \in A} (1-\lambda) \sum_{b \in B} \bar{z}_{q,z}(b) \tilde{V}_{\lambda,T}^1(\mu^+(\mu, a, b, z, q), q^+(b, z, q)). \quad (101)$$

Comparing equation (100) and (101) with equation (90) and (87), we see that the approximated security strategy of player 1 in discounted type 2 dual game is the security strategy of player 1 at stage 1 in  $T + 1$ -stage discounted type 2 dual game, and the approximated security strategy of player 2 in discounted type 1 dual game is the security strategy of player 2 at stage 1 in  $T + 1$ -stage discounted type 1 dual game. Following the same steps as in the proof of Theorem 4, we provide LP formulations to compute the approximated security strategies  $\tilde{\sigma}^\dagger$  and  $\tilde{\tau}^\dagger$  in the following theorem.

**Theorem 14.** Consider a  $T + 1$ -stage  $\lambda$ -discounted dual game  $\tilde{\Gamma}_{\lambda,T+1}^2(p, \nu)$ . Its game value  $\tilde{V}_{\lambda,T+1}^2(p, \nu)$  satisfies

$$\tilde{V}_{\lambda,T+1}^2(p, \nu) = \max_{x \in X, u_{:,0;\lambda,T+1} \in U, u_{:,0;\lambda,T+1} \in \mathbb{R}^{|L|}, \tilde{u} \in \mathbb{R}} \tilde{u} \quad (102)$$

$$s.t. \nu + u_{:,0;\lambda,T+1} \geq \tilde{u} \mathbf{1} \quad (103)$$

$$\lambda \sum_{k \in K} M^{klT} x_{k,h_1^A,h_1^B} + (1 - \lambda) u_{l,h_1^A,h_1^B;\lambda,T+1} T \mathbf{1} \geq u_{l,0;\lambda,T+1} \mathbf{1}, \quad \forall l \in L, \quad (104)$$

$$\lambda \sum_{k \in K} M^{klT} x_{k,h_{t+1}^A,h_{t+1}^B} + (1 - \lambda) u_{l,h_{t+1}^A,h_{t+1}^B;\lambda,T+1} T \mathbf{1} \geq u_{l,h_t^A,h_t^B;\lambda,T+1}^{a_t,b_t} \mathbf{1}, \quad (105)$$

$$\forall t = 1, \dots, T, l \in L, h_{t+1} \in H_{t+1}^A, h_{t+1}^B \in H_{t+1}^B,$$

where  $X$  is a set including all real vectors satisfying (3-4) with  $x_{:,h_0^A,h_0^B}^{a_0} = p$ , and  $U$  is an appropriately dimensional real space. The approximated security strategy  $\tilde{\sigma}^\dagger(\cdot, p, \nu)$  of player 1 in the discounted type 2 dual game  $\tilde{\Gamma}_\lambda^2(p, \nu)$  is his security strategy at stage 1 in the  $T + 1$ -stage discounted type 2 dual game  $\tilde{\Gamma}_{\lambda,T+1}^2(p, \nu)$ , and is computed in the following way

$$\tilde{\sigma}^\dagger(k, p, \nu) = \frac{x_{k,h_1^A,h_1^B}^*}{p^k}, \forall k \in K. \quad (106)$$

Similarly, the game value  $\tilde{V}_{\lambda,T+1}^1(\mu, q)$  of a  $T + 1$ -stage discounted type 1 dual game  $\tilde{\Gamma}_{\lambda,T+1}^1(\mu, q)$  satisfies

$$\tilde{V}_{\lambda,T+1}^1(\mu, q) = \min_{y \in Y, w_{:,0;\lambda,T+1} \in W, w_{:,0;\lambda,T+1} \in \mathbb{R}^{|K|}, \tilde{w} \in \mathbb{R}} \tilde{w} \quad (107)$$

$$s.t. \mu + w_{:,0;\lambda,T+1} \leq \tilde{w} \mathbf{1} \quad (108)$$

$$\lambda \sum_{l \in L} M^{kl} y_{l,h_1^A,h_1^B} + (1 - \lambda) w_{k,h_1^A,h_1^B;\lambda,T+1} \mathbf{1} \leq w_{k,0;\lambda,T+1} \mathbf{1}, \quad \forall k \in K, \quad (109)$$

$$\lambda \sum_{l \in L} M^{kl} y_{l,h_{t+1}^A,h_{t+1}^B} + (1 - \lambda) w_{k,h_{t+1}^A,h_{t+1}^B;\lambda,T+1} \mathbf{1} \leq w_{k,h_t^A,h_t^B;\lambda,T+1}^{a_t,b_t} \mathbf{1}, \quad (110)$$

$$\forall t = 1, \dots, T, k \in K, h_{t+1} \in H_{t+1}^A, h_{t+1}^B \in H_{t+1}^B,$$

where  $Y$  is a set including all real vectors satisfying (5-6) with  $y_{:,h_0^A,h_0^B}^{b_0} = q$ , and  $W$  is an appropriately dimensional real space. The approximated security strategy  $\tilde{\tau}^\dagger(\cdot, \mu, q)$  of player 2 in the discounted type 1 dual game  $\tilde{\Gamma}_\lambda^1(\mu, q)$  is his security strategy at stage 1 in  $T + 1$ -stage discounted type 1 dual game  $\tilde{\Gamma}_{\lambda,T+1}^1(\mu, q)$ , and is computed in the following way

$$\tilde{\tau}^\dagger(l, \mu, q) = \frac{y_{l,h_1^A,h_1^B}^*}{q^l}, \forall l \in L. \quad (111)$$

Corollary 9 says that player 1 and 2's security strategies in discounted primal game  $\Gamma_\lambda(p, q)$  are their security strategies in dual game  $\tilde{\Gamma}_\lambda^2(p, \nu^*)$  and  $\tilde{\Gamma}_\lambda^1(\mu^*, q)$ , respectively, where  $\mu^*$  and  $\nu^*$  are the solutions to the optimal problems on the right hand side of (64) and (66). Now, we know how to construct LP formulations to approximate the initial regrets, and players' security strategies in the dual games. We give the algorithms to compute the approximated security strategies for both players as below.

**Algorithm 15.** Player 1's approximated security strategy in discounted game  $\Gamma_\lambda(p, q)$

1) Initialization

- Set  $T$ , and read parameters:  $k$ ,  $M$ ,  $p$  and  $q$ .
- Given  $(p, q)$ , compute  $u_{:,0;\lambda,T}^*$  according to the LP (77-79).
- Set  $t = 1$ ,  $p_1 = p$  and  $\nu_1 = -u_{:,0;\lambda,T}^*$ .

2) Compute player 1's approximated security strategy  $\tilde{\sigma}^\dagger(\cdot, p_t, \nu_t)$  according to (106) based on the LP (102-105) with  $p = p_t$  and  $\nu = \nu_t$ .

3) Choose an action in  $A$  according to  $\tilde{\sigma}^\dagger(k, p_t, \nu_t)$ , and announce it publicly. Read player 2's action  $b_t$ .

4) Update  $p_{t+1}$  and  $\nu_{t+1}$  according to (49) and (72), respectively.

5) Update  $t = t + 1$  and go to step 2.

**Algorithm 16.** Player 2's approximated security strategy in discounted game  $\Gamma_\lambda(p, q)$



1) *Initialization*

- Set  $T$ , and read parameters:  $l$ ,  $M$ ,  $p$  and  $q$ .
- Given  $(p, q)$ , compute  $w_{:,0;\lambda,T}^*$  according to the LP (80-82).
- Set  $t = 1$ ,  $q_1 = q$  and  $\mu_1 = -w_{:,0;\lambda,T}^*$ .

2) Compute player 2's approximated security strategy  $\tilde{\tau}^\dagger(:, \mu_t, q_t)$  according to (111) based on the LP (107-110) with  $\mu = \mu_t$  and  $q = q_t$ .

3) Choose an action in  $B$  according to  $\tilde{\tau}^\dagger(l, \mu_t, q_t)$ , and announce it to the public. Read player 1's action  $a_t$ .

4) Update  $q_{t+1}$  and  $\mu_{t+1}$  according to (44) and (70), respectively.

5) Update  $t = t + 1$  and go to step 2.

*D. Performance analysis of the approximated security strategies*

With player 1 and 2's approximated security strategies  $\tilde{\sigma}^\dagger$  and  $\tilde{\tau}^\dagger$  described in Algorithm 15 and 16, we are interested in their worst case payoffs  $J^{\tilde{\sigma}^\dagger}$  and  $J^{\tilde{\tau}^\dagger}$ . Given player 1's strategy  $\sigma \in \Sigma$ , its worst case payoff in discounted game  $\Gamma_\lambda(p, q)$  is defined as

$$J^\sigma(p, q) = \min_{\tau \in \mathcal{T}} \gamma_\lambda(p, q, \sigma, \tau). \quad (112)$$

Similarly, given player 2's strategy  $\tau \in \mathcal{T}$ , its worst case payoff in discounted game  $\Gamma_\lambda(p, q)$  is defined as

$$J^\tau(p, q) = \max_{\sigma \in \Sigma} \gamma_\lambda(p, q, \sigma, \tau). \quad (113)$$

Because players' approximated security strategies in game  $\Gamma_\lambda(p, q)$  are derived from the approximated security strategies in its dual games, their worst case payoffs in  $\Gamma_\lambda(p, q)$  are highly related to the worst case payoffs in the dual games. We define player 1's worst case payoff  $\tilde{J}^{2,\sigma}$  in dual game  $\tilde{\Gamma}_\lambda^2(p, \nu)$  and player 2's worst case payoff  $\tilde{J}^{1,\tau}$  in dual game  $\tilde{\Gamma}_\lambda^1(\mu, q)$  as

$$\tilde{J}^{2,\sigma}(p, \nu) = \min_{q \in \Delta(L)} \min_{\tau \in \mathcal{T}} \tilde{\gamma}_\lambda^2(p, \nu, q, \sigma, \tau), \quad (114)$$

$$\tilde{J}^{1,\tau}(\mu, q) = \max_{p \in \Delta(K)} \max_{\sigma \in \Sigma} \tilde{\gamma}_\lambda^1(\mu, q, p, \sigma, \tau). \quad (115)$$

Following the same steps as in the proof of (63-66) in [12], [9], we can show the relations between  $J^\sigma(p, q)$  and  $\tilde{J}^{2,\sigma}(p, \nu)$ , and between  $J^\tau(p, q)$  and  $\tilde{J}^{1,\tau}(\mu, q)$  as below.

$$\tilde{J}^{2,\sigma}(p, \nu) = \min_{q \in \Delta(L)} \{J^\sigma(p, q) + q^T \nu\}, \quad (116)$$

$$J^\sigma(p, q) = \max_{\nu \in \mathbb{R}^{|L|}} \{\tilde{J}^{2,\sigma}(p, \nu) - q^T \nu\}, \quad (117)$$

$$\tilde{J}^{1,\tau}(\mu, q) = \max_{p \in \Delta(K)} \{J^\tau(p, q) + p^T \mu\}, \quad (118)$$

$$J^\tau(p, q) = \min_{\mu \in \mathbb{R}^{|K|}} \{\tilde{J}^{1,\tau}(\mu, q) - p^T \mu\}. \quad (119)$$

The worst case payoffs  $\tilde{J}^{2,\sigma}$  and  $\tilde{J}^{1,\tau}$  satisfy recursive formulas if  $\sigma$  and  $\tau$  are stationary strategies, i.e.  $\sigma$  only depends on  $p_t$  and  $\nu_t$ , and  $\tau$  only depends on  $\mu_t$  and  $q_t$ .

**Lemma 17.** *Let  $\sigma$  be player 1's stationary strategy that depends only on  $p_t$  and  $\nu_t$  in the discounted type 2 dual game  $\tilde{\Gamma}_\lambda^2(p, \nu)$ . Its worst case payoff  $\tilde{J}^{2,\sigma}(p, \nu)$  satisfies*

$$\tilde{J}^{2,\sigma}(p, \nu) = \tilde{F}_{\sigma(\cdot, p, \nu)}^{2, \tilde{J}^{2,\sigma}}(p, \nu). \quad (120)$$

*Similarly, let  $\tau$  be player 2's stationary strategy that depends only on  $\mu_t$  and  $q_t$  in the discounted type 1 dual game  $\tilde{\Gamma}_\lambda^1(\mu, q)$ . Its worst case payoff  $\tilde{J}^{1,\tau}(\mu, q)$  satisfies*

$$\tilde{J}^{1,\tau}(\mu, q) = \tilde{F}_{\tau(\cdot, \mu, q)}^{1, \tilde{J}^{1,\tau}}(\mu, q). \quad (121)$$

*Proof.* According to Bellman's principle, we have

$$J^\sigma(p, q) = \min_{z \in \Delta(B)^{|L|}} \left( \lambda \sum_{l \in L, k \in K} p^k q^l r^{kT} M^{kl} z^l + (1 - \lambda) \sum_{a \in A, b \in B} \bar{r}_{p,r}(a) \bar{z}_{q,z}(b) J^\sigma(p^+(a, p, r), q^+(b, q, z)) \right),$$

where  $r^k = \sigma(k, p, \nu)$ . From equation (116), we derive that

$$\begin{aligned} \tilde{J}^{2,\sigma}(p, \nu) &= \min_{q \in \Delta(L), z \in \Delta(B)^{|L|}} \left( \sum_{b \in B, l \in L} q^l z^l(b) \nu^l + \lambda \sum_{l \in L, k \in K, a \in A, b \in B} p^k q^l r^k(a) M_{ab}^{kl} z^l(b) \right. \\ &\quad \left. + (1 - \lambda) \sum_{a \in A, b \in B} \bar{r}_{p,r}(a) \bar{z}_{q,z}(b) J^\sigma(p^+(a, p, r), q^+(b, q, z)) \right) \\ &= (1 - \lambda) \min_{q \in \Delta(L), z \in \Delta(B)^{|L|}} \sum_{b \in B} \bar{z}_{q,z}(b) \sum_{a \in A} \bar{r}_{p,r}(a) \left( \sum_{l \in L} q^{+l}(b, q, z) \frac{\nu^l + \lambda \sum_{k \in K} p^{+k}(a, p, r) M_{ab}^{kl}}{1 - \lambda} \right. \\ &\quad \left. + J^\sigma(p^+(a, p, r), q^+(b, q, z)) \right). \end{aligned}$$

Since  $q^l z^l(b) = q^{+l}(b, q, z) \bar{z}_{q,z}(b)$  for any  $l \in L$  and  $b \in B$ , the minimum function taken with respect to  $q \in \Delta(L), z \in \Delta(B)^{|L|}$  is the same as the minimum function taken with respect to  $q^+ \in \Delta(L)^{|B|}, \bar{z} \in \Delta(B)$ . Thus, we have

$$\begin{aligned} \tilde{J}^{2,\sigma}(p, \nu) &= (1 - \lambda) \min_{q^+ \in \Delta(L)^{|B|}, \bar{z} \in \Delta(B)} \sum_{b \in B} \bar{z}(b) \sum_{a \in A} \bar{r}_{p,r}(a) \left( \sum_{l \in L} q^{+l}(b) \frac{\nu^l + \lambda \sum_{k \in K} p^{+k}(a, p, r) M_{ab}^{kl}}{1 - \lambda} \right. \\ &\quad \left. + J^\sigma(p^+(a, p, r), q^+(b)) \right) \\ &= (1 - \lambda) \min_{\bar{z} \in \Delta(B)} \sum_{b \in B} \bar{z}(b) \sum_{a \in A} \bar{r}_{p,r}(a) \tilde{J}^{2,\sigma}(p^+(a, p, r), \nu^+(a, b, r, p)) \\ &= (1 - \lambda) \min_{b \in B} \sum_{a \in A} \bar{r}_{p,r}(a) \tilde{J}^{2,\sigma}(p^+(a, p, r), \nu^+(a, b, r, p)) = \tilde{F}_{\sigma(\cdot, p, \nu)}^{2, \tilde{J}^{2,\sigma}}(p, \nu). \end{aligned}$$

The second equality is derived from (116).

Following the same steps, we can show that  $\tilde{J}^{1,\tau}(\mu, q) = \tilde{F}_{\tau(\cdot, \mu, q)}^{1, \tilde{J}^{1,\tau}}(\mu, q)$ .  $\square$

Based on Lemma 17, we are ready to analyze the performance difference between players' approximated security strategies and their security strategies.

**Theorem 18.** Consider a discounted game  $\Gamma_\lambda(p, q)$ . If player 2 uses  $\tilde{\sigma}^\dagger$  defined in (100) as his strategy, and follows Algorithm 15 to take actions, then his worst case payoff  $J^{\tilde{\sigma}^\dagger}(p, q)$  satisfies

$$\|J^{\tilde{\sigma}^\dagger} - V_\lambda\|_{\sup} \leq \frac{2(1 - \lambda)}{\lambda} \|V_{\lambda|T} - V_\lambda\|_{\sup}. \quad (122)$$

If player 2 uses  $\tilde{\tau}^\dagger$  defined in (101) as his strategy, and follows Algorithm 16 to take actions, then his worst case payoff  $J^{\tilde{\tau}^\dagger}(p, q)$  satisfies

$$\|J^{\tilde{\tau}^\dagger} - V_\lambda\|_{\sup} \leq \frac{2(1 - \lambda)}{\lambda} \|V_{\lambda|T} - V_\lambda\|_{\sup}. \quad (123)$$

*Proof.* According to equation (117) and (64), we have  $|J^{\tilde{\sigma}^\dagger}(p, q) - V_\lambda(p, q)| = |\max_{\nu \in \mathbb{R}^{|L|}} \{\tilde{J}^{2, \tilde{\sigma}^\dagger}(p, \nu) - q^T \nu\} - \max_{\nu \in \mathbb{R}^{|L|}} \{\tilde{V}_\lambda^2(p, \nu) - q^T \nu\}|$ . Let  $\nu^*$  be the solution to the optimal problem  $\max_{\nu \in \mathbb{R}^{|L|}} \{\tilde{V}_\lambda^2(p, \nu) - q^T \nu\}$ . Since  $J^{\tilde{\sigma}^\dagger}(p, q) \leq V_\lambda(p, q)$ , we have

$$|J^{\tilde{\sigma}^\dagger}(p, q) - V_\lambda(p, q)| \leq |\tilde{J}^{2, \tilde{\sigma}^\dagger}(p, \nu^*) - \tilde{V}_\lambda^2(p, \nu^*)| \leq \|\tilde{J}^{2, \tilde{\sigma}^\dagger} - \tilde{V}_\lambda^2\|_{\sup}, \forall p \in \Delta(K), q \in \Delta(L) \quad (124)$$

According to equation (120), (99) and (97), we have for any  $p \in \Delta(K)$  and any  $\nu \in \mathbb{R}^{|L|}$ ,

$$\begin{aligned} |\tilde{J}^{2, \tilde{\sigma}^\dagger}(p, \nu) - \tilde{V}_\lambda^2(p, \nu)| &\leq |\tilde{J}^{2, \tilde{\sigma}^\dagger}(p, \nu) - \tilde{V}_{\lambda, T+1}^2(p, \nu)| + |\tilde{V}_{\lambda, T+1}^2(p, \nu) - \tilde{V}_\lambda^2(p, \nu)| \\ &\leq |\tilde{F}_{\tilde{\sigma}^\dagger(\cdot, p, \nu)}^{2, \tilde{J}^{2, \tilde{\sigma}^\dagger}}(p, \nu) - \tilde{F}_{\tilde{\sigma}^\dagger(\cdot, p, \nu)}^{2, \tilde{V}_{\lambda, T}^2}(p, \nu)| + (1 - \lambda) \|\tilde{V}_{\lambda, T}^2 - \tilde{V}_\lambda^2\|_{\sup} \\ &\leq (1 - \lambda) \|\tilde{J}^{2, \tilde{\sigma}^\dagger} - \tilde{V}_{\lambda, T}^2\|_{\sup} + (1 - \lambda) \|\tilde{V}_{\lambda, T}^2 - \tilde{V}_\lambda^2\|_{\sup}. \end{aligned}$$

Thus, we have  $\|\tilde{J}^{2, \tilde{\sigma}^\dagger} - \tilde{V}_\lambda^2\|_{\sup} \leq (1 - \lambda) \|\tilde{J}^{2, \tilde{\sigma}^\dagger} - \tilde{V}_{\lambda, T}^2\|_{\sup} + (1 - \lambda) \|\tilde{V}_{\lambda, T}^2 - \tilde{V}_\lambda^2\|_{\sup} \leq (1 - \lambda) \|\tilde{J}^{2, \tilde{\sigma}^\dagger} - \tilde{V}_\lambda^2\|_{\sup} + 2(1 - \lambda) \|\tilde{V}_{\lambda, T}^2 - \tilde{V}_\lambda^2\|_{\sup}$ , which implies that

$$\|\tilde{J}^{2, \tilde{\sigma}^\dagger} - \tilde{V}_\lambda^2\|_{\sup} \leq \frac{2(1 - \lambda)}{\lambda} \|\tilde{V}_{\lambda, T}^2 - \tilde{V}_\lambda^2\|_{\sup}.$$

After applying the above inequality to (124), we have for any  $p \in \Delta(K)$  and  $q \in \Delta(L)$ ,  $|J^{\tilde{\sigma}^\dagger}(p, q) - V_\lambda(p, q)| \leq \frac{2(1 - \lambda)}{\lambda} \|\tilde{V}_{\lambda, T}^2 - \tilde{V}_\lambda^2\|_{\sup}$ , which implies that  $\|J^{\tilde{\sigma}^\dagger} - V_\lambda\|_{\sup} \leq \frac{2(1 - \lambda)}{\lambda} \|\tilde{V}_{\lambda, T}^2 - \tilde{V}_\lambda^2\|_{\sup}$ . According to equation (93), equation (122) is shown.

With the same technique, equation (123) can be shown to be true.  $\square$

TABLE I  
TOTAL CHANNEL CAPACITY

$\begin{matrix} \backslash \\ k \end{matrix} \begin{matrix} I \\ \end{matrix}$	1(0.5 km)		2 (2 km)	
1 ([1 1] km)	108.89	113.78	122.30	154.40
	108.89	113.78	122.30	154.40
2 ([1 5] km)	11.48	107.38	24.89	107.42
	99.04	20.15	100.26	60.77
3 ([5 5] km)	1.64	13.75	2.85	13.79
	1.64	13.75	2.85	13.79

#### IV. CASE STUDY: JAMMING IN UNDERWATER SENSOR NETWORKS

The jamming in underwater sensor networks is originally modelled as a two-player zero-sum one-shot Bayesian game in [14]. We adopt the game model in [14], and extend it to a repeated Bayesian game with uncertainties on both the sensors' positions and the jammer's position.

Let's assume that there are two sensors in the network which send data to a sink node through a shared spectrum at [10, 40] kHz. The distance from a sensor to the sink node is either 1 km or 5 km. The shared spectrum is divided into two channels,  $\mathcal{B}_1 = [10, 25]$  kHz and  $\mathcal{B}_2 = [25, 40]$  kHz. Generally speaking, channel 1 works much better for a sensor far away, and almost the same as channel 2 for a sensor close by. The sensors need to coordinate with each other to use the two channels to transfer as much data as possible to the sink node in the presence of a jammer. The jammer's distance from the sink node is 0.5 km or 2 km. While the jammer doesn't know the sensors' positions, the sensors don't know the jammer's position either. For every time period, the jammer can only generate noises in one channel, which can be detected by the sensors. At the same time, the jammer can also observe whether a channel is used by a far-away sensor or a close-by sensor. The jammer's goal is to minimize the data transmitted through the two channels.

The sensors (player 1) have three types according to their position distribution, which are [1 1] (type 1), [1 5] (type 2), and [5, 5] (type 3). We consider [1 5] and [5 1] as one type. The initial distribution over the three types is  $p_0 = [0.5 \ 0.3 \ 0.2]$ . When playing the game, they have two choices, sensor 1 uses channel 1 while sensor 2 uses channel 2 (action 1) or sensor 1 uses channel 2 while sensor 2 uses channel 1 (action 2). The jammer (player 2) has two types according to his position, which are 0.5 (type 1) and 2 (type 2), and the initial distribution over the two types is  $q_0 = [0.5 \ 0.5]$ . His actions are jamming channel 1 (action 1) or channel 2 (action 2). Suppose both the sensors and the jammer transmit with constant power 95 dB re  $\mu$ Pa. A channel's capacity can be computed based on the Shannon-Hartley theorem with the average under water signal-to-noise ratio described in [15], [14]. The payoff matrices, whose element is the total channel capacity measured by bit/s given both players' types and actions, are given in Table I.

We first consider a two-stage Bayesian repeated game between the sensors and the jammer. Based on the linear program (11-13), we compute the sensors' security strategy shown in Table II with a security level to be 162.49 bit/s. According to the linear program (15-17), the jammer's security strategy is computed, and given in Table III. The jammer's security level is 162.49 bit/s which meets the sensors' security level. We then use the players' security strategies in Table II and II in the two-stage under water jamming game. The jamming game was run for 100 times for each experiment, and we did the experiment for 30 times. The total channel capacity in the jamming game varies from 142.12 bit/s to 185.23 bit/s with an average capacity to be 162.79 bit/s, which is very close to the game value computed according to (11-13) and (15-17).

Next, we would like to use security strategies based on fixed-sized sufficient statistics in the jamming game, and see whether we can still achieve the game value. First of all, we need to verify Theorem 5. According to Lemma 3 and linear program (15-17) and (11-13), the initial regret  $\mu^*$  in type 2 dual game  $\tilde{\Gamma}_T^2(p_0, \mu^*)$  is  $[-145.45 \ -179.53]$ , and the initial regret  $\nu^*$  in type 1 dual game  $\tilde{\Gamma}_T^1(\nu^*, q_0)$  is  $[-234.77 \ -141.44 \ -13.38]$ . Player 1's security strategy in dual game  $\tilde{\Gamma}_T^2(p_0, \mu^*)$  is computed according to the linear program (53-55), and given in Table IV. We see that player 1's security strategy in dual game  $\tilde{\Gamma}_T^2(p_0, \mu^*)$  is different from but very close to player 1's security strategy in the primal game  $\Gamma_T(p_0, q_0)$ . The security level of  $\tilde{\sigma}^*$  in the primal game is 162.49 (checked by building a linear program the same as (11-13) with  $x$  fixed), the game value of the primal game. Therefore,  $\tilde{\sigma}^*$  is player 1's another security strategy. Player 2's security strategy in the dual game  $\tilde{\Gamma}_T^1(\nu^*, q_0)$  is computed according to linear program (58-61), and given in Table V, which matches player 2's security strategy in the primal game  $\Gamma_T(p_0, q_0)$ . We then run the two-stage under water jamming game using security strategies based on fixed sized sufficient statistics, and followed Algorithm 6 and 7 to take actions. For each experiment, the two-stage under water jamming game was run for 100 times, and we did 30 experiments. The channel capacity varies from 144.38 to 180.31 bit/s, with an average capacity to be 162.32 bit/s, which is almost the same as the game value 162.49 bit/s.

Finally, we test Algorithm 15 and 16 in the discounted under water jamming game with discount constant  $\lambda = 0.7$  to see whether the outcome satisfies our anticipation. In the algorithms, we set  $T = 3$ , and  $V_{\lambda,3} = 78.28$  bit/s. First, we found that the highest game value of a 3-stage discounted game occurs at  $p_0 = [1 \ 0 \ 0]$  and  $q_0 = [0 \ 1]$ , and  $\|V_{\lambda,3}\|_{\sup} = 118.99$  bit/s. Second, we found an upper bound on  $\|V_{\lambda}(p_0, q_0)\|_{\sup}$ . According to equation (93) and (98), we have  $\|V_{\lambda} - V_{\lambda,3}\|_{\sup} \leq (1 - \lambda)^3 \|V_{\lambda}\|_{\sup}$ , which implies that  $\|V_{\lambda}\|_{\sup} \leq 1/(1 - (1 - \lambda)^3) \|V_{\lambda,3}\|_{\sup} = 122.29$  bit/s. Third, we derive a lower bound

TABLE II  
 $\sigma_t^{1*}(k, h_t^A, h_t^B)$  IN  $\Gamma_T(p_0, q_0)$

$k \backslash h_t^A, h_t^B$	$\emptyset, \emptyset$	1,1	1,2	2,1	2,2
1	0.43	0.5	0.5	0.5	0.5
2	0.18	0	0.23	0	0.39
3	0.44	0.5	0.5	0.5	0.5

TABLE III  
 $\tau_t^{1*}(l, h_t^A, h_t^B)$  IN  $\Gamma_T(p_0, q_0)$

$l \backslash h_t^A, h_t^B$	$\emptyset, \emptyset$	1,1	1,2	2,1	2,2
1	0.068	0	0.5	0	0.5
2	1	1	\	1	\

TABLE IV  
 $\tilde{\sigma}_t^{1*}(k, h_t^A, h_t^B)$  IN  $\tilde{\Gamma}_T^2(p_0, \mu^*)$

$k \backslash h_t^A, h_t^B$	$\emptyset, \emptyset$	1,1	1,2	2,1	2,2
1	0.33	0.5	0.5	0.5	0.5
2	0.18	0	0.25	0.068	0.38
3	0.45	0.5	0.5	0.5	0.5

TABLE V  
 $\tilde{\tau}_t^{1*}(l, h_t^A, h_t^B)$  IN  $\tilde{\Gamma}_T^1(\nu^*, q_0)$

$l \backslash h_t^A, h_t^B$	$\emptyset, \emptyset$	1,1	1,2	2,1	2,2
1	0.068	0	0.5	0	0.5
2	1	1	\	1	\

on the security level of the sensors' approximated security strategy. According to equation (122), (93) and (99), we have  $J^{\tilde{\sigma}^\dagger}(p_0, q_0) \geq V_{\lambda,3}(p_0, q_0) - 2(1 - \lambda)^4 / \lambda \|V_\lambda\|_{\sup} \geq 75.44$  bit/s. Finally, we get an upper bound on the security level of the jammer's approximated security strategy. According to equation (123), (93) and (98), we have  $J^{\tilde{\tau}^\dagger}(p_0, q_0) \leq V_\lambda(p_0, q_0) + 2(1 - \lambda)^4 / \lambda \|V_\lambda\|_{\sup} \leq V_{\lambda,3}(p_0, q_0) + (1 - \lambda)^3 \sum_{t=1}^{\infty} \lambda(1 - \lambda)^{t-1} 154.4 + 2(1 - \lambda)^4 / \lambda \|V_\lambda\|_{\sup} \leq 85.28$  bit/s. Therefore, our anticipated channel capacity in the discounted under water jamming game is between 75.44 and 85.28 bit/s. Now, we run the discounted under water jamming game (10 stages) for 100 times. For each run, we truncate the infinite horizon discounted game to 10 stages, since the total channel capacity for the truncated stages is less than  $10^{-3}$  bit/s. The average channel capacity is 82.15 bit/s, which is within our anticipation, and verifies our main results in the discounted games.

## V. FUTURE WORK

This paper studies two-player zero-sum repeated Bayesian games, and provides LP formulations to compute players' security strategies in finite horizon case and approximated security strategies in discounted infinite horizon case. In both cases, strategies based on fixed-sized sufficient statistic are provided. The fixed-sized sufficient statistics for each player consists of the belief over his own type and the regret state with respect to the other player's type. We are interested in extending the results to two player non-zero-sum stochastic Bayesian games in the future.

## REFERENCES

- [1] D. Rosenberg, "Duality and markovian strategies," *International Journal of Game Theory*, vol. 27, no. 4, pp. 577–597, 1998.
- [2] A. Nayyar, A. Gupta, C. Langbort, and T. Başar, "Common information based markov perfect equilibria for stochastic games with asymmetric information: Finite games," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 555–570, 2014.
- [3] Y. Ouyang, H. Tavafoghi, and D. Teneketzis, "Dynamic games with asymmetric information: Common information based perfect bayesian equilibria and sequential decomposition," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 222–237, 2017.
- [4] D. Vasal and A. Anastasopoulos, "A systematic process for evaluating structured perfect bayesian equilibria in dynamic games with asymmetric information," in *American Control Conference (ACC)*, 2016. IEEE, 2016, pp. 3378–3385.
- [5] D. Fudenberg and J. Tirole, "Perfect bayesian equilibrium and sequential equilibrium," *Journal of Economic Theory*, vol. 53, no. 2, pp. 236–260, 1991.
- [6] R. J. Aumann and M. Maschler, *Repeated games with incomplete information*. MIT press, 1995.
- [7] S. Zamir, "Repeated games of incomplete information: Zero-sum," *Handbook of Game Theory*, vol. 1, pp. 109–154, 1992.
- [8] T. Başar and G. J. Olsder, *Dynamic noncooperative game theory*. SIAM, 1998.
- [9] S. Sorin, *A first course on zero-sum repeated games*. Springer Science & Business Media, 2002, vol. 37.
- [10] V. S. Kamble, "Games with vector payoffs: a dynamic programming approach," Ph.D. dissertation, University of California, Berkeley, 2015.
- [11] B. Von Stengel, "Efficient computation of behavior strategies," *Games and Economic Behavior*, vol. 14, no. 2, pp. 220–246, 1996.
- [12] B. De Meyer, "Repeated games and partial differential equations," *Mathematics of Operations Research*, vol. 21, no. 1, pp. 209–236, 1996.
- [13] T. Sandholm, "The state of solving large incomplete-information games, and application to poker," *AI Magazine*, vol. 31, no. 4, pp. 13–32, 2010.

- [14] V. Vadori, M. Scalabrin, A. V. Guglielmi, and L. Badia, "Jamming in underwater sensor networks as a bayesian zero-sum game with position uncertainty," in *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, pp. 1–6.
- [15] N. Baldo, P. Casari, and M. Zorzi, "Cognitive spectrum access for underwater acoustic communications," in *ICC Workshops-2008 IEEE International Conference on Communications Workshops*. IEEE, 2008, pp. 518–523.